



UNIVERSIDAD
DE MÁLAGA

Facultad de Ciencias

Departamento de Biología Celular, Genética y Fisiología

TESIS DOCTORAL

**Variabilidad genética de cuasiespecies de virus RNA y
ssDNA *in vivo* ante cambios en la tasa de error.**

Luis Díaz Martínez

Directores:

Ana Grande Pérez

Enrique Viguera Mínguez



Málaga 2016



UNIVERSIDAD
DE MÁLAGA

AUTOR: Luis Díaz Martínez

 <http://orcid.org/0000-0002-7659-4349>

EDITA: Publicaciones y Divulgación Científica. Universidad de Málaga



Esta obra está bajo una licencia de Creative Commons Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional:

<http://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>

Cualquier parte de esta obra se puede reproducir sin autorización
pero con el reconocimiento y atribución de los autores.

No se puede hacer uso comercial de la obra y no se puede alterar, transformar o hacer
obras derivadas.

Esta Tesis Doctoral está depositada en el Repositorio Institucional de la Universidad de
Málaga (RIUMA): riuma.uma.es



La Dra. Ana Grande Pérez, profesora Titular del Departamento de Biología Celular, Genética y Fisiología de la Universidad de Málaga,

INFORMA:

Que el trabajo de Tesis Doctoral titulado “Variabilidad genética de cuasiespecies de virus RNA y ssDNA *in vivo* ante cambios en la tasa de error” presentada por el Licenciado Luis Díaz Martínez para optar al grado de Doctor en Biotecnología Avanzada, se ha realizado en el departamento de Biología Celular, Genética y Fisiología de la Facultad de Ciencias, bajo mi dirección y, considerando que se ha concluido, autorizo su presentación para que pueda ser evaluada por el tribunal correspondiente.

Y para que así conste y tenga los efectos que correspondan en cumplimiento de la legislación vigente, firmo el presente informe.

Fdo: Ana Grande Pérez

Málaga, Junio 2016



El Dr. Enrique Viguera Mínguez, profesor Titular del Departamento de Biología Celular, Genética y Fisiología de la Universidad de Málaga,

INFORMA:

Que el trabajo de Tesis Doctoral titulado “Variabilidad genética de cuasiespecies de virus RNA y ssDNA *in vivo* ante cambios en la tasa de error” presentada por el Licenciado Luis Díaz Martínez para optar al grado de Doctor en Biotecnología Avanzada, se ha realizado en el departamento de Biología Celular, Genética y Fisiología de la Facultad de Ciencias, bajo mi dirección y, considerando que se ha concluido, autorizo su presentación para que pueda ser evaluada por el tribunal correspondiente.

Y para que así conste y tenga los efectos que correspondan en cumplimiento de la legislación vigente, firmo el presente informe.

Fdo: Enrique Viguera Mínguez

Málaga, Junio 2016



D. Francisco Javier Ruiz Albert, Coordinador del Máster Universitario en Biotecnología Avanzada de la Universidad de Málaga,

AUTORIZA:

la presentación de la Tesis Doctoral titulada “Variabilidad genética de cuasiespecies de virus RNA y ssDNA *in vivo* ante cambios en la tasa de error”, presentada por el Licenciado Luis Díaz Martínez para optar al grado de Doctor en Biotecnología Avanzada.

Y para que así conste y tenga los efectos que correspondan en cumplimiento de la legislación vigente, firmo el presente informe.

Fdo: Francisco Javier Ruiz Albert.

Málaga, Junio 2016

TRIBUNAL DE EVALUACIÓN DE LA TESIS DOCTORAL:

Presidente

Dr. Eduardo Rodríguez Bejarano

Universidad de Málaga (UMA)

Málaga

Secretario

Dr. Esteban Domingo Soláns

Centro de Biología Molecular "Severo Ochoa" (CBMSO)

Madrid

Vocal

Dr. Luis Blanco Dávila

Centro de Biología Molecular "Severo Ochoa" (CBMSO)

Madrid

Suplentes

Dra. Celia Perales Viejo

Instituto de Investigación Valle de Hebrón

Barcelona

Dr. Rafael Rodríguez Ariza

Universidad de Córdoba (UCO)

Córdoba

Dr. Guillermo Thode Mayoral

Universidad de Málaga (UMA)

Málaga

El trabajo recogido en la siguiente memoria ha sido llevado a cabo en el Departamento de Biología Celular, Genética y Fisiología de la Facultad de Ciencias de la Universidad de Málaga, bajo la dirección de los doctores Ana grande Pérez y Enrique Viguera Mínguez.

El trabajo de Investigación ha sido financiado por los proyectos P09-CVI-5428 y P10-CVI-6561 de la Consejería de Economía y Conocimiento de la Junta de Andalucía.

Luis Díaz Martínez ha disfrutado de una beca-contrato de Excelencia de la Junta de Andalucía asociado al proyecto P10-CVI-6561 de la Consejería de Economía y Conocimiento de la Junta de Andalucía.

Parte de los resultados de esta Tesis Doctoral han sido presentados en varios congresos Nacionales e Internacionales:

XVI Congreso Nacional de la Sociedad Española de Fitopatología (SEF). 2012. Málaga. Oral en sesión paralela.

XII Congreso Nacional de la Sociedad Española de Virología (SEV). 2013. Burgos. Oral en sesión plenaria.

XVI International Congress on Molecular Plant-Microbe Interactions (IS-MPMI). 2014. Rodas, Grecia. Póster.

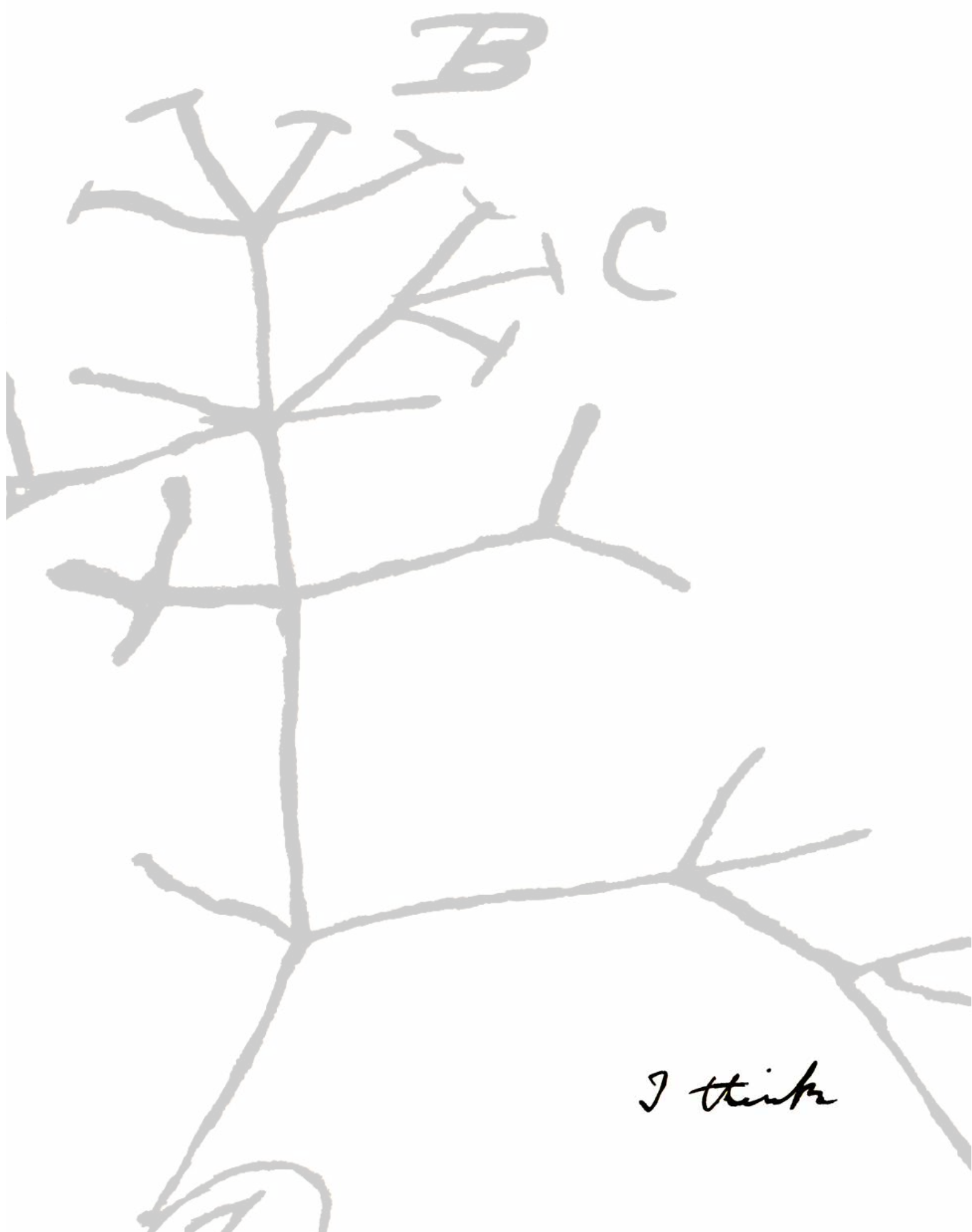
XII Congreso Nacional de la Sociedad Española de Virología (SEV). 2015. Madrid. Póster.

NGS 2016 *Genome anotation Congress*. 2016. Barcelona. Póster

A mi familia

A Laura

Darío...



I think

Página anterior: Composición realizada a partir del libro de notas de Charles Darwin, donde plasmó la idea de que todos los organismos comparten un ancestro común, dibujando el primer árbol filogenético conocido en la historia de la ciencia.

AGRADECIMIENTOS

Este trabajo que hoy presento es el resultado de un esfuerzo colectivo, tanto científico como afectivo. Han sido cuatro maravillosos años en los que he podido ver mi crecimiento como persona y como profesional en la ciencia. Debo mucho a muchas personas, y procuraré pagar más de lo debido.

No puedo comenzar los agradecimientos sin decirle a **Ana Grande** GRACIAS, pero gracias de verdad. Necesitaría mucho más que este pequeño espacio para agradecerte todo lo que has hecho por mí. Gracias por darme una oportunidad que no esperaba tener nunca, por confiar en mí y darme la libertad que necesitaba dentro del laboratorio, por tus consejos en mis momentos de dudas y agobios, por tu infinita paciencia y por darme a conocer el maravilloso mundo de las cuasiespecies. Quiero que seas consciente de cómo has cambiado mi vida con esta oportunidad, y quiero que sepas que lo mucho o poco que consiga profesionalmente habrá sido gracias a ti.

Y **Enrique Viguera**, tu tampoco te vas a escapar. Quiero agradecerte inmensamente que me hayas enseñado el camino que debe seguir un buen científico, ser exacto en sus definiciones, trabajador y ser un friki de lo que haces. Quiero agradecerte que siempre hayas tenido la puerta de tu “torre de marfil” abierta para mí y para mis dudas. Quiero agradecerte tu entusiasmo por cada nuevo resultado, por querer ir siempre un paso más allá, me encanta esa visión aplicada que tienes de todo lo que hacemos en el grupo.

Necesito también daros un gracias repartido entre ambos por “este último esfuerzo”, me habéis demostrado que soy algo más que vuestro doctorando. Aquí me tendréis siempre que me necesitéis.

Ana María, “jefa”, gracias por la confianza que depositaste en mí desde el primer día que llegué al Zaidín de estancia. Gracias por ser una amiga y preocuparte por mí, aun cuando ya estaba de vuelta por tierras malagueñas...Gracias por ser lo mejor de Granada.

¿Y a vosotros que os digo panda de individuos? A los que aún estais a mi lado, Alba, Adri, Eloy, Diego, Nieves, Javi, Jose, Pablo...(por dios que no se me haya olvidado nadie) mil gracias por aguantarme cada día por hacer de estos cuatro años en Málaga una experiencia inolvidable.

Diego ya sabes que tengo debilidad contigo...aquí tienes un amigo para toda la vida, aunque tú no quieras. Eres una de las personas más nobles y trabajadoras que he conocido, no cambies nunca. Bueno, si trabajas un poquito menos tampoco te va a pasar nada...

Josico, ya te lo he dicho mil veces, eres mi Macho, mi Alberto Macho particular, gracias por ser tan paciente con cualquier duda que te planteaba, eres la generosidad personificada y no sabes como me alegro de que la vida te haya puesto en mi camino.

Eloy, sé que esperas mucho de lo que pueda decirte aquí, gracias.

Pablo, yo no se ya ni qué decirte, eres el “amo del calabozo”. Mil gracias por mantener siempre esa actitud tan positiva y echarte a la espalda más de lo que deberías. Quiero que sepas que el laboratorio es mejor desde que tú estas aquí, no tengas ninguna duda.

Adri, pequeño padawan del iphone laser, gracias por ser mi media naranja friki, no sabes cuanto necesitaba encontrar alguien con quien hablar de estas cosas.

Alba, gracias por tu alegría, por tus payasadas, por tu humor tan blanco como tu forma de ser. Eres digna sucesora de mi asiento.

Y a mis dos últimas mejores incorporaciones **Nieves y Javi**, vais a ser lo que queráis ser, sois geniales, gracias por aguantar mis idas de olla, es que tengo predilección por sacarle los colores a los pekes del grupo. Me teneis para los que necesitéis, no solo para el Nanovue...

Eloy, “miarma” que no me olvido de ti. Gracias por hacer que en me sintiese en Granada como en casa. ¡Eres el siguiente!

Pedro S., te debo la mitad de la tesis, espero que esta experiencia juntos te haya aportado tanto como a mí. Gracias por traducir a código lo que sólo podía imaginar en mi cabeza.

A los que ya no estais en el labo pero que compartisteis gran parte del camino que hoy termino de recorrer: **Pi y Ale**, cuanto me gustaría teneros más cerquita. Gracias por hacernos sentir a Laura, a mí y al pequeño Darío como parte de vuestra familia, sois personas de luz.

Isa P., “Aragonchi”, **Isa Brichette**: mis Isas, gracias por todo lo que me habéis enseñado, si soy un poquitito mejor que cuando empecé es gracias a vosotras.

No puedo ni quiero olvidarme de los “exportados” **Ainhoa** ¿escribiré bien tu nombre alguna vez?, “**Charo**”, Mi “**Peperoni**” de mi cuore, **Juanjo y Adeli**, mil gracias por hacer que me integrase en el grupo, a pesar de lo asocial que pudiera parecer en un principio al final no lo era tanto ¿no?. **Axier**, gracias por tu buen hacer en el laboratorio, eres el ejemplo perfecto de alumno interno.

También tengo que agradecer a todos los seniors que me han hecho crecer en lo científico: **Miguel Ángel Botella**, te debo mucho más de lo que crees. **Sr. Rodríguez Bejarano**, gracias por ser mi “Geminiciclopedia”, no puede haber nadie que sepa más de nuestros “amigos” que tú. Ha sido una suerte poder disfrutar de tus conocimientos. **Araceli**, tengo que agradecerte la implicación que pones en cada seminario, cada semana, de cada mes, de cada año. Cada vez que me toca exponer siento como si fueses la persona más atenta de la sala, sabes sacar lo mejor de nosotros, eres inspiradora. **Javi y Carmen**, ha sido una alegría cruzarme con vosotros cada día en el laboratorio, irradiáis alegría y buenas vibraciones. **Cayo**, gracias por ser el adalid de las causas perdidas a favor del doctorando, hacen falta más personas como tú. **Edgar**, tú y tus refranes se merecen estar entre los grandes, gracias por compartir tanto, ojalá volviéramos a coincidir.

A los “torrecitos” **Blanca, Pepe, Alvaro Piedra y Melillero, Jessi, José, Manolo, Sonia, Migue...** gracias por compartir grandes momentos de seminarios y celebraciones, me llevo un trocito de cada uno. A ti **Ali** te pongo solita porque quiero que sepas que no tendré tiempo suficiente para devolverte la paciencia que tuviste conmigo cuando daba mis primeros pasos. Mil gracias.

No me puedo olvidar de mis niños de La Rábida, especialmente de **Diego, Jose, Eli, Laurita, Ana “la yanqui”, mi Verito** (que poquito te queda para leer). Sobran las palabras, habrá distancia entre nosotros pero siempre os llevaré en mi corazón. **Julio y Happy Eli** gracias por ser parte de mi familia, por compartir y ayudar tanto cuando os he necesitado.

A mi Familia, **Papá**, te debo la mitad de lo que soy, gracias por animarme siempre a esforzarme un poco más. Por tu tiempo invertido en mí. Gracias por cada día que viste amanecer en tu trabajo. Gracias por ser un hombre recto y con ideales. Espero haber podido aprender algo de ti en todo este tiempo. Nunca, nunca, podré devolverte todo lo que me has dado. **Hermana**, muchas gracias por estar pendiente de nosotros, por tenernos siempre presente aun estando tan lejos. Gracias por convertirte en la amiga en que te has convertido. Te has hecho una mujer y no me he dado ni cuenta. Aun me queda mucho que disfrutar de ti. A mis tíos **Manoli y Pepe** por los buenos momentos pero sobre todo por ser sostén en los malos momentos. Nunca podré olvidar lo que hicisteis por nosotros. Gracias. Al resto de mi familia, abuelos, tíos y primos. A la Familia que elegimos, **Fabi, Juan Luis, Mari Carmen, Antonio, Filo, Sandra**, gracias por todo lo que me habéis dado en la vida.

Mazagoneros, **Jorge, Migue, Ana, Aída, Diana, Inma, Esther, Jorge O., mis Cordobeses**, a todos, gracias por ser mi remanso de paz. Sois mi infancia, mi juventud y sereis parte de mi vejez.

A mis suegros **Lauro y Dafro**, gracias por acogerme como a un hijo. Sólo puedo daros las gracias una y otra vez por todo lo que nos dais diariamente.

A mi **Sebas**, mi profesora de Biología del Instituto. Gracias por despertar en mí el gusanillo de la Ciencia, estoy aquí gracias a ti.

Al amor de mi vida, **Laura**, gracias por compartir este viaje conmigo, contra viento y marea. Gracias por demostrarme que soy lo primero para ti. Gracias por abrazar mi mitad triste y convertirla en una alegría diaria, por comprender mis rarezas y aceptarlas tal y como son. Gracias por haberme elegido. No va a bastarme una vida para devolverte tanto amor y entrega como tu me has dado. Gracias por darme lo mejor de mi vida, nuestro pequeño **Darío**. Entre los dos habéis llenado mi vida de amor y felicidad. Os amo cada día más.

Y a ti, **Mamá**, por la vida y porque sé que allí donde estés, hoy estarás orgullosa de tu hijo. Siempre te querré.

GRACIAS.

LUIS DÍAZ MARTÍNEZ
Junio de 2016

INDICE

ÍNDICE

ABREVIATURAS	1
Abreviaturas de virus	3
Otras abreviaturas.....	4
RESUMEN.....	7
INTRODUCCIÓN GENERAL	11
I. Los virus como patógenos de plantas.....	13
II. Variabilidad genética en las poblaciones virales.....	16
III. Fuentes de variación genética	20
a. Mutación	20
b. Recombinación	22
c. Intercambio de componentes genómicos	24
d. Segmentación	24
OBJETIVOS	25
CAPÍTULO 1	29
Introducción.....	31
1. Catástrofe de error, defeción letal y mutagénesis letal como estrategia antiviral.....	31
2. TMV como modelo de estudio de cuasiespecies virales y mutagénesis letal	36
3. TMV y el género tobamovirus	37
3.1. Estructura y función de las regiones no codificantes de TMV	38
3.2. Estructura y función de las proteínas virales.....	39
3.3. Estructura del virión y ciclo de vida	39
3.4. Mutagénesis química de TMV	42
3.5. Estructura y metabolismo del 5-FU	44
3.6. Sistema modelo TMV- <i>N. tabacum</i> para el estudio de la mutagénesis letal	47
Materiales y métodos.....	49
1. Material biológico	49
1.1. Plantas	49
1.2. Cultivos celulares	50
1.3. Cepas bacterianas.....	50
1.4. Stock de TMV.....	50
1.5. Extractos de TMV.....	51
2. Medios de cultivo.....	52
2.1. Medios de cultivo bacteriano	52
2.2. Medios para cultivo de <i>Nicotiana tabacum</i>	52
2.3. Análogo de base	53
3. Cultivo y crecimiento de organismos	53
3.1. Cultivo y crecimiento de bacterias	53
3.2. Cultivo y crecimiento de plantas	53
3.3. Mantenimiento de los callos de BY-2	54
3.4. Ensayo de lesiones locales.....	55
4. Oligonucleótidos	55
4.1. Cebadores empleados para la retrotranscripción de TMV.....	56
4.2. Cebadores empleados para RT-qPCR	56
4.3. Otros cebadores	56

5. Extracción de nucleótidos y ácidos nucleicos	57
5.1. Extracción de nucleótidos para análisis de HPLC.....	57
5.2. Extracción de RNA total de <i>Nicotiana tabacum</i> cv Samsun nn.....	58
6. Tratamiento enzimático de ácidos nucleicos	58
6.1. Tratamiento con DNasa	58
6.2. Clonaje de fragmentos de DNA a vectores plasmídicos	59
7. Amplificación de ácidos nucleicos	59
7.1. PCR estándar.....	59
7.2. Síntesis de cDNA mediante transcripción inversa	60
7.3. PCR Touchdown.....	60
7.4. RT-PCR cuantitativa (RT-qPCR)	61
7.5. Cuantificación absoluta	63
7.6. Amplificación por círculo rodante (RCA)	64
8. Electroforesis de ácidos nucleicos	64
8.1. Electroforesis en geles de agarosa	64
8.2. Purificación de los fragmentos de DNA de geles de agarosa.....	65
9. Transformación genética.....	65
9.1. Preparación de células competentes.....	65
9.2. Clonaje molecular para el análisis genético de poblaciones de TMV	65
9.3. Infección de <i>N. tabacum</i>	65
10. Análisis de niveles de nucleótidos intracelulares	66
11. Secuenciación	66
12. Caracterización de la variabilidad genética de las cuasiespecies.	67
12.1. Frecuencias de mutación mínima y máxima	67
12.2. Diversidad nucleotídica	67
12.3. Índice de Shannon normalizado	68
13. Estadística.....	68
14. Bioinformática	68
Resultados	71
1. Toxicidad del análogo de base 5-FU en plantas de <i>Nicotiana tabacum</i> cv Samsun	71
2. Efecto del 5-FU sobre la infectividad y la acumulación de TMV.....	71
2.1. Infectividad	72
2.2. Determinación de la carga vírica	73
2.3. Determinación de la infectividad específica	74
3. Caracterización genética de los espectros de mutantes	76
4. Complejidad de las cuasiespecies de TMV sometidas a tratamiento con 5-FU	77
5. Heterogeneidad de las cuasiespecies víricas.....	81
6. Caracterización de las mutaciones en presencia de 5-FU	81
7. Análisis del espacio de secuencia	84
8. Dirección de las fuerzas selectivas.....	86
9. Influencia del FU en la distribución de la variabilidad de TMV	88
10. Análisis filogenético del espectro de mutantes.....	89
11. Determinación de ribonucleótidos trifosfato (NTPs) y FUTP intracelulares	93
Discusión.....	97
Anexos	103
CAPÍTULO 2	109
Introducción.....	111
1. Plataformas actuales de secuenciación masiva.....	112
2. Aplicaciones de la secuenciación masiva en Virología.....	113

3.	La secuenciación masiva en el estudio de la diversidad viral.....	114
4.	Software bioinformático para análisis de datos procedentes de NGS	116
Materiales y Métodos		119
1.	Material biológico.....	119
1.1.	Plantas	119
2.	Extracción y purificación de ácidos nucleicos.....	119
2.1.	Extracción de DNA de <i>S. lycopersicum</i> infectadas con TYLCV.....	119
3.	Tratamiento enzimático de ácidos nucleicos	119
3.1.	Tratamiento con endonucleasas de restricción.....	119
3.2.	Eliminación de endonucleasas de las muestras.....	120
4.	Reacciones de amplificación de ácidos nucleicos	120
4.1.	Amplificación por círculo rodante (RCA)	120
5.	Electroforesis de ácidos nucleicos	121
5.1.	Electroforesis en geles de agarosa	121
6.	Secuenciación	121
7.	Bioinformática	125
7.1.	Equipo informático	125
7.2.	Programas.....	126
7.3.	Análisis estadísticos	129
Resultados		131
1.	Análisis bioinformático de secuencias de virus mediante la herramienta QuasiFlow. .	131
1.1.	Descripción	131
1.2.	Desarrollo	131
2.	Aplicación de QuasiFlow en análisis experimentales	143
2.1.	Identificación de moléculas recombinantes en coinfecciones de TYLCV	143
2.2.	Caracterización de variantes de nucleótidos con baja frecuencia en DNA mitocondrial humano	149
Discusión.....		169
CAPITULO 3		173
Introducción.....		175
1.	Origen de la variabilidad genética de los virus de ssDNA	175
2.	Polimerasas de DNA.....	177
3.	Síntesis de translesión en eucariotas.....	180
3.1.	Familia Y.....	181
4.	DNA polimerasas TLS descritas en plantas y animales	182
4.1.	DNA polymerasa η	182
4.2.	DNA polymerase κ	183
4.3.	Rev1	183
4.4.	PrimPol una nueva primasa polimerasa implicada en TLS	184
5.	TYLCV y el género <i>Begomovirus</i>	187
5.1.	Estructura y función de la región intergénica no codificante	189
5.2.	Estructura y función de las proteínas virales.....	190
5.3.	Estructura del virión y ciclo de vida	192
5.4.	TYLCV como modelo de estudio de cuasiespecies virales de ssDNA <i>in vivo</i>	194
Materiales y Métodos		197
1.	Material biológico	197
1.1.	Plantas	197
1.2.	Cepas bacterianas.....	197

1.3. Clon infectivo.....	198
2. Medios de cultivo.....	199
2.1. Medios de cultivo bacteriano	199
2.2. Medios para cultivo de <i>Arabidopsis thaliana</i>	199
3. Cultivo y crecimiento de organismos	200
3.1. Cultivo y crecimiento de bacterias	200
3.2. Cultivo y crecimiento de plantas	200
4. Oligonucleótidos	203
4.1. Cebadores para el genotipado de las líneas mutantes de <i>A. thaliana</i>	203
4.2. Cebadores para la cuantificación relativa de mRNA de <i>Arabidopsis thaliana</i>	205
4.3. Cebadores para la cuantificación absoluta de la carga viral de TYLCV	206
5. Extracción y purificación de ácidos nucleicos.....	207
5.1. Purificación de DNA plasmídico.....	207
5.2. Extracción de DNA de <i>Arabidopsis thaliana</i> infectadas con TYLCV	207
5.3. Extracción de RNA total de <i>Arabidopsis thaliana</i>	207
6. Tratamiento enzimático de ácidos nucleicos	208
6.1. Tratamiento con endonucleasas de restricción.....	208
6.2. Eliminación de endonucleasas de las muestras.....	208
7. Reacciones de amplificación de ácidos nucleicos	209
7.1. PCR estándar.....	209
7.2. PCR para genotipado de plantas.....	209
7.3. Síntesis de DNAc mediante transcripción inversa	210
7.4. PCR cuantitativa.....	211
7.5. Amplificación por círculo rodante (RCA)	212
8. Electroforesis de ácidos nucleicos	213
8.1. Electroforesis en geles de agarosa	213
9. Transformación genética.....	214
9.1. Preparación de células competentes.....	214
9.2. Transformación de <i>Escherichia coli</i>	214
9.3. Transformación de <i>Agrobacterium tumefaciens</i>	215
9.4. Agroinfección de <i>A. thaliana</i>	215
10. Secuenciación	215
11. Bioinformática	216
11.1. Uso del programa QuasiFlow.....	216
11.2. Programas para la gestión de datos y análisis de secuencias.....	217
Resultados	219
1. Caracterización de las líneas mutantes de <i>A. thaliana atrev1-1, atrev1-2, atpolh, atpolk</i> y <i>atprimpol</i>	219
1.1. Fenotipado de las líneas mutantes de <i>Arabidopsis thaliana</i>	220
1.2. Genotipado de las líneas mutantes de <i>Arabidopsis thaliana</i>	221
1.3. Expresión relativa de las diferentes DNA polimerasas TLS sobre fondos genéticos mutantes de <i>Arabidopsis thaliana</i>	224
2. Cinética de la carga viral de TYLCV sobre fondos genéticos mutantes de DNA polimerasas TLS.....	226
3. Caracterización de las cuasiespecies crecidas en ausencia de DNA polimerasas TLS....	229
4. Caracterización del espacio de secuencia y variabilidad genética de las cuasiespecies virales de TYLCV en ausencia de REV1 y PRIMPOL	234
4.1. Análisis de mutaciones de TYLCV en ausencia de Rev1 y PrimPol.....	236
4.2. Análisis de InDels de TYLCV en ausencia de Rev1 y PrimPol	240
4.3. Análisis de eventos de recombinación en las cuasiespecies virales de TYLCV en ausencia de Rev1 y PrimPol.....	243

5. Reconstrucción de haplotipos en las cuasiespecies virales de TYLCV en ausencia de Rev1 y PrimPol.....	246
6. Reconstrucción de la cuasiespecie viral de TYLCV en ausencia de Rev1 y PrimPol	246
7. Análisis multivariable de las infecciones de TYLCV sobre fondos genéticos mutantes de <i>Arabidopsis thaliana</i>	248
Discusión.....	251
CONCLUSIONES	259
BIBLIOGRAFÍA	263

ABREVIATURAS

ABREVIATURAS

Abreviaturas de virus

CaMV	<i>Cauliflower mosaic virus</i>
CLCuV	<i>Cotton leaf curl virus</i>
EBV	<i>Epstein-Barr virus</i>
FBNSV	<i>Faba bean necrotic stunt virus</i>
FMDV	<i>Foot-and-mouth disease virus</i>
HCMV	<i>Human cytomegalovirus</i>
HCV	<i>Hepatitis c virus</i>
HIV	<i>Human immunodeficiency virus</i>
HSV-1	<i>Herpes simplex virus tipo 1</i>
LCMV	<i>Lymphocytic choriomeningitis virus</i>
MSV	<i>Maize streak virus</i>
MSV	<i>Maize streak virus</i>
MuLV	<i>Murine leukemia virus</i>
TEV	<i>Tobacco etch virus</i>
TMV	<i>Tobacco mosaic virus</i>
TYLCAxV	<i>Tomato yellow leaf curl Axarquia virus</i>
TYLCCNV	<i>Tomato yellow leaf curl China virus</i>
TYLCMaV	<i>Tomato yellow leaf curl Malaga virus</i>
TYLCSV	<i>Tomato yellow leaf curl Sardinia virus</i>
TYLCV	<i>Tomato yellow leaf curl virus</i>
TYLCV-IL	<i>Tomato yellow leaf curl virus Israel</i>
TYLCV-Mld	<i>Tomato yellow leaf curl virus Mild</i>
VSV	<i>Vesicular stomatitis virus</i>

Otras abreviaturas

°C	Grados Celsius
5-FU	5-fluorouracilo
A	Adenina
BY-2	Bright yellow-2
C	Citosina
CP	Proteína de la cápside
CSR	Replicación de la cadena complementaria
CV	Coeficiente de variación
d	Distancia genética media
dATP	Deoxiadenosín trifosfato
dCTP	Deoxicitosín trifosfato
dGTP	Deoxiguanosin trifosfato
DNA	Acido desoxirribonucleico
DNAc	DNA complementario
DNAmt	DNA mitocondrial
<i>dns</i>	Diversidad de sustituciones no sinónimas por sitio no sinónimo
dpi	Días postinfección
<i>ds</i>	Diversidad de sustituciones sinónimas por sitio sinónimo
dsDNA	DNA de cadena doble
dTMP	Desoxitimidina monofosfato
dUMP	Deoxiuridina monofosfato
dUTP	Deoxiuridin trifosfato
FdUMP	Fluorodeoxiuridina monofosfato
FdUTP	Fluorodeoxiuridina trifosfato
FUDR	Fluorodesoxiuridina
FUMP	Fluorouridina monofosfato
FUTP	Fluorouridina trifosfato
G	Guanina
g	Gramos
HEL/MET	Metiltransferasa/helicasa
HR	Hipersensibilidad
ICTV	Comité Internacional de Taxonomía de Virus
InDel	Inserción/delección
IPTG	Isopropil-β-D-1-tiogalactopiranósido
kDa	KiloDalton
l	Litros
lfu	Unidades formadoras de lesión
mg	Miligramos
min	Minutos

ml	Mililitros
mM	Milimolar
MP	Proteína del movimiento
mut	Mutación
NGS	Secuenciación de nueva generación
NHEJ	Unión de extremos no homólogos
NSP	<i>Nuclear shuttle protein</i>
nt	Nucleótido(s)
ORF	Marco abierto de lectura
p	Peso
PCA	Análisis de componentes principales
PCR	Reacción en cadena de la polimerasa
RCA	Amplificación por círculo rodante
RCR	Replicación por círculo rodante
RdDp	RNA polimerasas dependientes de DNA
RDR	Replicación dependiente de recombinación
RdRp	RNA polimerasas dependientes de RNA
rep	Replicación
RNA	Ácido ribonucleico
RNA_m	RNA mensajero
RNA_{nc}	RNA no codificante
RNA_r	RNA ribosómico
RNA_{sg}	RNA subgenómico
RNA_v	RNA viral
rpm	Revoluciones por minuto
RTs	Reversotranscriptasas
s	Segundos
Sn	Índice de Shannon normalizado
SNPs	Polimorfismo de un nucleótido
ssDNA	DNA de cadena sencilla
ssRNA	RNA de cadena sencilla
T	Timina
TCA	Ácido tricloroacético
T_m	Temperatura de fusión
TMD	Enfermedad del mosaico del tabaco
TOA	Trioctilamina
ts	Sensible a temperatura
TS	Timidilato sintasa
T_s	Transiciones
TU	Tiouracilo
T_v	Trasversiones
TYLCD	Enfermedad del rizado amarillo del tomate
U	Uracilo

UDG	Uracil-DNA-glicosilasa
UK	Uridina quinasa
UP	Uridina fosforilasa
UTR	Región no traducida
v	Volumen
VRC	Complejos de replicación virales
W	Vatios
µg	Microgramos
µl	Microlitros

RESUMEN

El objetivo general de esta tesis es entender la estructura poblacional *in vivo* de las cuasiespecies de virus de RNA y ssDNA cuando se altera su variabilidad genética debido a cambios en la tasa de error. El abordaje de este objetivo general lo hemos llevado a cabo desde dos aproximaciones diferentes. Por un lado, hemos estudiado las bases moleculares de la mutagénesis letal *in vivo* por tratamiento de un virus RNA de plantas, el virus del mosaico del tabaco (TMV) con un agente mutagénico. Para ello infectamos plantas de *Nicotiana tabacum* crecidas *in vitro* y las tratamos durante 10 días en presencia de diferentes concentraciones del análogo de base 5-fluorouracilo (5-FU). Tras el tratamiento cultivamos las plantas 21 días sin análogo *ex vitro*. Las poblaciones de RNA genómico de TMV se cuantificaron por RT-qPCR y se determinó su infectividad por ensayo de lesiones locales. Además se secuenció por Sanger un amplicón conteniendo parte del ORF que codifica la RNA polimerasa (RdRp) y el ORF de la proteína MP para obtener la secuencia consenso y, tras su clonaje, secuencias de clones moleculares del espectro de mutantes. Los resultados mostraron que el 5-FU reduce la infectividad de TMV pero no la acumulación de moléculas genómicas lo que disminuye su infectividad específica y que esta es recuperada tras 21 días sin análogo. El análisis de la complejidad y heterogeneidad así como el tipo de mutaciones predominantes, indica que el efecto antiviral del análogo está más relacionado con el efecto del mismo sobre la distribución de la variabilidad genética originada por la acción el 5-FU y que esta actúa de diferente manera en las dos regiones analizadas. Este trabajo apoya el modelo de defección letal para la mutagénesis letal *in vivo* de TMV y abre las puertas a emplear este sistema de virus RNA en plantas para el estudio *in vivo* de las bases moleculares de la mutagénesis letal mediante agentes mutagénicos.

En segundo lugar de esta tesis hemos desarrollado una herramienta bioinformática llamada QuasiFlow para poder analizar simultáneamente y de manera sencilla, robusta y rápida la variabilidad de las cuasiespecies virales a partir de datos de NGS. Esta herramienta nos ha permitido reconstruir haplotipos completos, detectar genomas virales recombinantes y variantes genómicas minoritarias presentes en los espectros de mutantes de geminivirus en coinfecciones de *Tomato yellow leaf curl virus*-Mld, *Tomato yellow leaf curl virus*-IL, *Tomato yellow leaf curl Malaga virus* en tomate. Además, hemos podido demostrar la versatilidad de Quasiflow al extender su utilización a un estudio sobre variantes genómicas minoritarias en muestras de genomas DNA mitocondrial.

Por último, hemos abordado el estudio del origen de la variabilidad genética de un virus de DNA de cadena sencilla, el geminivirus TYLCV, mediante las nuevas tecnologías de secuenciación masiva en combinación con el análisis de QuasiFlow. Estos virus de ssDNA se caracterizan por tener una variabilidad genómica similar a la de virus RNA. Sin embargo, sus genomas carecen de genes codificantes para una DNA polimerasa replicativa, por lo que esta variabilidad debe ser originada por polimerasas celulares. En la célula, además de DNA polimerasas replicativas, se han identificado DNA polimerasas de tolerancia al daño o de síntesis de translesión con altas tasas de error. En esta Tesis hemos estudiado el papel de las DNA polimerasas de la Familia Y de *Arabidopsis thaliana* en la variabilidad de las cuasiespecies de TYLCV. Nuestros resultados apoyan la hipótesis de que dado que el DNA monocatenario es más susceptible al daño, estas polimerasas intervendrían en su reparación llevando a cabo una acción mutadora sobre él. Además, recientemente se ha descubierto en humanos una DNA primasa-polimerasa llamada PrimPol, con cierta actividad TLS que, conforme a nuestros resultados, podría estar implicada en las fases iniciales de la infección de TYLCV así como contribuir a la variabilidad genética de geminivirus interviniendo en la recombinación ilegítima de sus genomas.

INTRODUCCIÓN GENERAL

I. Los virus como patógenos de plantas

Los virus (del griego *lós*=veneno), junto con los viroides, son los organismos auto replicantes más pequeños que existen en la Tierra. En su versión más básica, consisten simplemente en un pequeño segmento de ácido nucleico envuelto en un caparazón de proteínas (Knipe et al., 2001). Se ha discutido mucho acerca de si los virus son organismos vivos o no (Forterre, 2010; Moreira and López-García, 2009), sin llegar a un acuerdo unánime acerca de esta cuestión. Lo que sí está claro es que los virus existen en dos estados diferentes. Cuando no están en contacto con la célula hospedadora, el virus permanece en estado latente. Durante este periodo dentro del virus no existe ningún tipo de actividad biológica, y en esencia el virus no es más que una partícula orgánica estática. En este estado tan simple de partícula “no-viva” los virus son denominados viriones. Estos viriones pueden permanecer en estado latente por largos periodos de tiempo, hasta entrar en contacto con el hospedador apropiado. Cuando esto sucede, el virión se vuelve activo y es a partir de esta fase cuando es referido como virus. En estas circunstancias los virus muestran todos los rasgos característicos de cualquier organismo vivo, como puede ser reaccionar a su entorno y dirigir todos sus esfuerzos hacia la autorreplicación (Bândeá, 1983).

En la **Figura 1** se ilustra la clasificación actual de los virus de plantas según el Comité Internacional de Taxonomía de Virus (*International Committee on Taxonomy of Viruses*, ICTV) en función de la naturaleza de su material genético. Este puede ser de DNA o de RNA, de cadena doble o sencilla, circular o lineal y, adicionalmente, el material genético puede presentarse como genomas multipartitos o fragmentados. Los genomas de RNA de cadena sencilla (*single stranded RNA*, ssRNA) pueden presentar orientación positiva, con función de RNA mensajero (RNAm) o negativa, en cuyo caso actuaría como secuencia molde del RNAm viral.

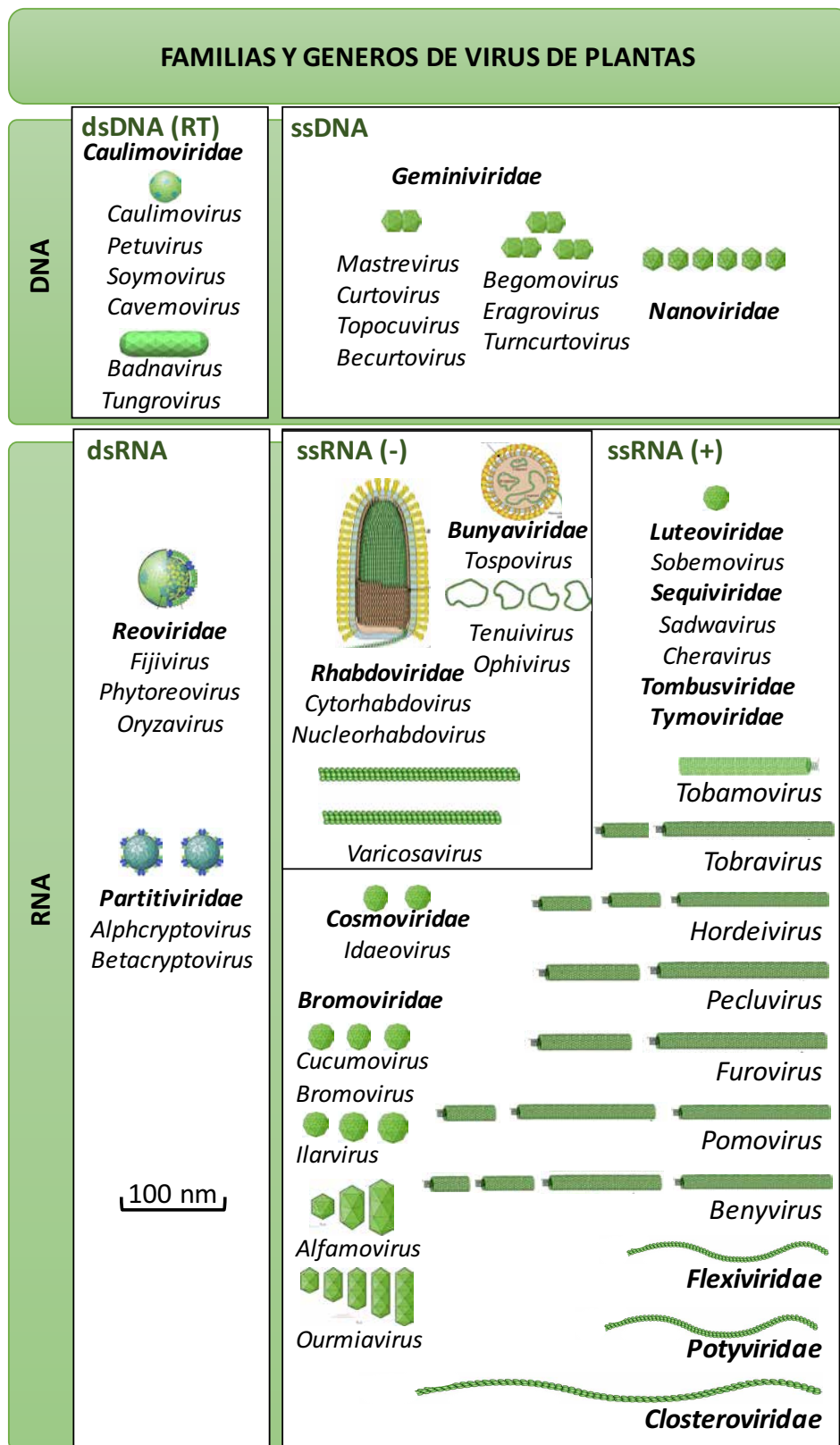


Figura 1. Clasificación de los virus de plantas. Adaptado del VIII Informe del Comité Internacional de Taxonomía de Virus, 2012. Las representaciones de los virus han sido extraídas con autorización de la página web www.viralzone.expasy.org

Desde finales del siglo XIX hasta hoy, las plantas han servido como modelo ideal para el estudio de infecciones virales. Las plantas son seres vivos sésiles, este hecho hace que los virus de plantas tengan que servirse de vectores de transmisión para pasar de un hospedador a otro. La etapa de transmisión en el ciclo de infección de cualquier virus de planta es un paso vital porque es el que asegura el mantenimiento, la supervivencia y la propagación del virus. El ciclo de transmisión en virus de plantas implica un proceso de adquisición del virus, cuando el vector se alimenta de una planta infectada, retención estable en el vector, transporte de las partículas virales dentro de este y posterior inoculación del virus en un nuevo hospedador durante una nueva toma de alimento por parte del vector. La mayoría de los virus de plantas (76%) son transmitidos por vectores como insectos, nematodos y hongos. Muchos de estos vectores suponen de por sí una plaga para muchos cultivos, convirtiéndolos en agentes ideales para una dispersión a medias y largas distancias (Hogenhout et al., 2008).

El proceso de invasión del huésped, implica múltiples etapas como la desencapsidación, la replicación, la transcripción, el movimiento, la supresión de las defensas de la célula huésped, etc. Todos estos pasos necesitan de múltiples interacciones del virus con numerosas factores celulares que contribuyen a la formación de complejos de ácidos nucleicos y proteínas. En el caso de los virus de plantas de RNA, la infección comienza con la asociación del RNA viral a la membrana del retículo endoplasmático o a membranas de orgánulos que estén conectados a este. La traducción de este RNA viral produce factores virales que interaccionan con ellos mismos y con proteínas de la célula formando los complejos de replicación virales (*Viral Replication Complexes*, VRC) (Ritzenthaler and Hofmann, 2007; Ueki and Citovsky, 2011), revisado en (Heinlein, 2015). Estos VRC interaccionan con el citoesqueleto de la célula permitiendo el movimiento de estos complejos a través de los plasmodesmos. Los mecanismos por los que los VRCs se asocian a membranas reflejan un uso deliberado de la actina y los microtúbulos por parte del virus (Niehl et al., 2013; Brandizzi and Wasteneys, 2013). Así, los procesos de replicación y movimiento de los virus están íntimamente relacionados y no se trata de eventos aislados. Debido a que las dimensiones de estos VRC superan el diámetro de poro de los plasmodesmos, los virus codifican proteínas de movimiento

(MP) para modificar este diámetro de poro y facilitar el paso (Fujiki et al., 2006), (Hammond-Kosack, 2000; Lucas, 2006).

En el caso de virus de ssDNA, como los geminivirus bipartitos, se ha descrito la interacción entre la proteína vírica *nuclear shuttle protein* (NSP) y la histona H3, permitiendo el desplazamiento del DNA viral como un minicromosoma. Además, estos minicromosomas se han detectado asociados a túbulos del retículo endoplasmático de la célula infectada (Ward et al., 1997). A diferencia de los virus de RNA, los virus de ssDNA necesitan replicarse en el núcleo de la célula y posteriormente provocar la salida de las moléculas virales de su interior. En este caso NSP también juega un papel fundamental al interaccionar con una GTPasa (NIG) asociada a la cara externa de la envuelta nuclear y que facilita el paso de las moléculas de virus a través de los poros nucleares (revisado en Hanley-Bowdoin et al., 2013).

El descubrimiento de cada nuevo factor de la planta involucrado en la infección informa sobre los mecanismos moleculares que subyacen tras la susceptibilidad y la expresión de síntomas del huésped. Algunos factores son necesarios para la compatibilidad, mientras que otros modulan la infección compitiendo por los recursos, degradando el RNA del virus (silenciamiento), apagando la expresión génica (*shut-off*), o mediando una respuesta de hipersensibilidad (HR) por genes de resistencia (Kang et al., 2005). Por otro lado, es evidente que todos y cada uno de los eventos celulares, dentro de cada etapa del ciclo infectivo, están fuertemente interconectados y coordinados. Esta alta interconexión invita a una visión más holística, donde la célula infectada por el virus debe percibirse como la unidad funcional de la infección (Ahlquist et al., 2003).

II. Variabilidad genética en las poblaciones virales

En la actualidad, la aparición o emergencia de nuevos virus de plantas representa una amenaza directa para la agricultura. Por eso es importante determinar y comprender los factores genéticos y evolutivos implicados en la aparición de nuevos virus.

Esta evolución tan rápida que se da en los virus es consecuencia de sus genomas compactos, elevadas tasas de mutación, ciclos replicativos cortos y grandes tamaños poblacionales que, en conjunto, generan poblaciones altamente variables. Todo ello hace de los virus un sistema ideal para estudiar las tasas de cambios evolutivos puesto

que su evolución es medible dentro de la escala de tiempo de la observación humana (Drummond et al., 2003; revisado en Duffy and Holmes, 2008).

El estudio de la variabilidad genética y la evolución de las poblaciones son importantes para distintas áreas de la Biología. En los virus de plantas, el conocimiento de su evolución es esencial para el desarrollo de estrategias de control, ya que en muchos casos, las metodologías de control no son efectivas debido a la diversidad genética de las poblaciones y a su capacidad de evolucionar rápidamente (García-Arenal and McDonald, 2007; Acosta-Leal et al., 2011). La capacidad de los virus y los viroides a cambiar rápidamente subyace en muchas de las preocupaciones a la hora de gestionar la enfermedad. Muchas veces el ascenso y la caída de los diferentes genotipos en una población viral se debe a los efectos de la selección natural: los genomas variantes que generan descendientes más viables se vuelven más frecuentes con el tiempo. Este proceso se puede acelerar o inhibir por los cuellos de botella que sufre un patógeno de planta al pasar de una célula a otra y de una planta a otra (French and Stenger, 2003; French and Stenger, 2005; Froissart et al., 2005). Los cuellos de botella poblacionales reducen drásticamente el tamaño poblacional, produciendo la amplificación de subpoblaciones de genomas que pueden diferir de la población parental en su composición y, en consecuencia, alterar las propiedades biológicas del conjunto (Escarmís et al., 1996), (Pfeiffer and Kirkegaard, 2006).

Para ampliar su gama de huéspedes, una población de virus ya debe contener alguna variante (tal vez en una frecuencia muy baja de la población) que pueda infectar al hospedador en potencia. En este nuevo hospedador, los mutantes estarán más adaptados y aumentarán en frecuencia, dando lugar a la emergencia de una nueva cepa capaz de infectar este nuevo hospedador. La comprensión de los procesos que generan la diversidad viral y los procesos ecológicos que determinan las presiones selectivas y cuellos de botella pueden indicar el camino de posibles intervenciones o determinar dónde y cuándo las medidas de control podrían ser más eficaces (Acosta-Leal et al., 2011)

Las poblaciones de virus de RNA tienen asociada una alta variabilidad y heterogeneidad genética, otorgándoles una gran capacidad adaptativa (revisado en Domingo et al., 2012) pero en los últimos años se ha demostrado que algunas poblaciones de virus de ssDNA

y de dsDNA también presentan estas características (Duffy and Holmes, 2008; Van Loy et al., 2013; Van Loy et al., 2015).

Desde que se tiene conocimiento científico de esta variabilidad en las poblaciones virales se han propuesto numerosos nombres para intentar caracterizar a una población viral altamente variable. Variación intrahospedador, diversidad intrahospedador, mezcla de mutantes, degeneración nucleotídica, hiperploidía, heteroespecie o heteropoblación son algunos de los términos empleados para definir a estas poblaciones virales altamente variables y heterogéneas (Domingo et al., 2012). De entre todos estos términos el de cuasiespecie viral es el que está arraigado con más fuerza y el que más se adapta a nuestra forma de entender la variabilidad de las poblaciones virales, es por tanto, el término al que se hará referencia a lo largo de esta Tesis Doctoral.

La teoría de cuasiespecies fue desarrollada en la década de los 70 por los científicos Manfred Eigen y Peter Schuster como un modelo de evolución molecular que permitía describir la estructura poblacional y la adaptabilidad de los replicones primitivos (Eigen, 1971; Eigen and Shuster, 1979). Los primeros modelos utilizados para la demostración de esta teoría fueron el fago Q β , el virus de la fiebre aftosa (FMDV) y el virus de la estomatitis vesicular (VSV) (Domingo et al., 1978; Domingo et al., 1980; Holland et al., 1982; Sobrino et al., 1986).

Las cuasiespecies víricas se definen como un conjunto de genomas virales no idénticos pero estrechamente relacionados entre sí, que están sometidos a cambios en sus secuencias, competición y selección, y cuya unidad de selección es la población viral (revisado por Domingo et al., 2012). La distribución de las variantes en las cuasiespecies no es arbitraria, sino que se centra en torno a una *secuencia maestra*, que define a la secuencia dominante en la población por tener mayor capacidad replicativa (*fitness*). Además de la secuencia maestra, toda cuasiespecie viral es caracterizada por su secuencia consenso, que es la secuencia nucleotídica que presenta en cada posición el nucleótido más frecuente en el conjunto de genomas de la población. Dicha secuencia no tiene por qué estar representada en el espectro de mutantes (Biebricher and Eigen, 2006) (**Figura 2**).

Las propiedades de una cuasiespecie dependen de los genomas que la forman y de cómo interaccionan, positiva (complementación) o negativamente (interferencia), los unos

con los otros. La gran variedad genotípica y fenotípica de los genomas que la componen dotan a la cuasiespecie de una alta capacidad de adaptación a distintos ambientes y circunstancias adversas. La existencia de multitud de genomas distintos hacen posible ampliar el rango del hospedador y variar el grado de virulencia (Baranowski et al., 2003; Bordería et al., 2010; Domingo et al., 2012). El tratamiento de enfermedades de origen vírico se hace difícil por la generación o selección de mutantes resistentes a los agentes antivirales en la población (Domingo et al., 2012).

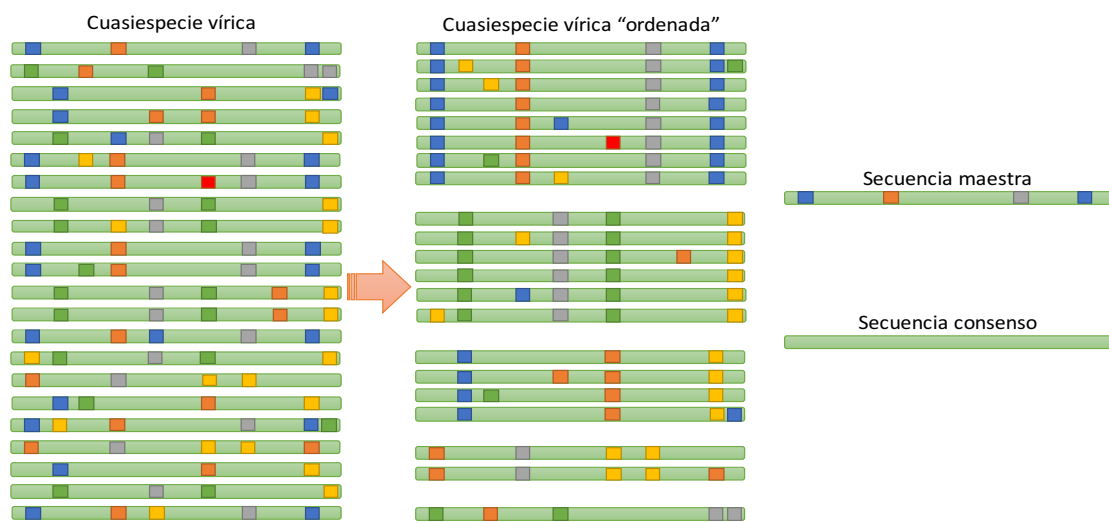


Figura 2. Distribución de una cuasiespecie vírica. La cuasiespecie está formada por un espectro de mutantes donde cada genoma de la población viral está representado por una barra verde y las mutaciones están representadas con cuadros de diferentes colores. Dentro de la población existe una secuencia maestra que es la más representada dentro del espectro de mutantes, y una secuencia consenso que es la media de todas las variantes dentro de la población vírica, pudiendo no existir dentro de la población. Figura adaptada de Domingo et al, 2012.

En una población de virus en evolución los mutantes están en continua competición y sometidos a la selección natural tanto positiva como negativa (Domingo et al., 2012). En las cuasiespecies virales, dado que sus componentes comparten el carácter seleccionable, lo que se selecciona es una distribución de en lugar de un solo individuo. En la selección positiva un conjunto de genotipos se hacen dominantes como resultado de la evaluación positiva de los caracteres fenotípicos expresados en ellos. En cambio cuando opera la selección negativa los mutantes relacionados con el carácter negativamente evaluado se eliminan o se mantienen a baja frecuencia en la

cuasiespecie. La distinción entre selección negativa o positiva puede ser algo confusa porque en ese proceso de competición entre mutantes del espectro, la selección negativa de determinados conjuntos de genotipos que son eliminados podría interpretarse como la selección positiva de aquellos mutantes que permanecen en la cuasiespecie.

III. Fuentes de variación genética

En los virus, los mecanismos que generan la variación genética son la mutación, la recombinación, el intercambio de componentes genómicos y la segmentación.

a. Mutación

La mutación es la fuente principal de variación en las poblaciones virales. Es un proceso por el cual se incorporan nucleótidos que no están presentes en la secuencia molde en la cadena de DNA que se está sintetizando. Este fenómeno incluye tanto mutaciones puntuales como inserciones o deleciones (InDels).

La tasa de mutación hace referencia al número de errores genéticos (mutaciones puntuales, inserciones y deleciones) que se acumulan por unidad de tiempo (Ge et al., 2007), por generación (Chao et al., 2002) o por ronda de replicación genómica (Drake, 1991; Drake, 1993; Drake et al., 1998;). Por otro lado, la tasa de sustitución nucleotídica se define como el número de cambios mutacionales fijados (por selección natural o por deriva genética) por sitio nucleotídico y por unidad de tiempo (usualmente, en años) (Duffy and Holmes, 2008). Finalmente, la frecuencia de mutación representa la proporción de nucleótidos mutados que observamos en una población de genomas (Domingo et al., 2006). La frecuencia de mutación viene expresada como mutaciones por nucleótido y está influida por múltiples factores ambientales y bioquímicos, siendo uno de los más mencionados tradicionalmente la fidelidad de la polimerasa viral, pero también por la competición, la selección y, en algunos casos, eventos aleatorios.

Existen varias estimas de las tasas de mutación viral que reflejan los cambios de secuencia nucleotídica que ocurren en cada ronda de replicación. Estas estimas abarcan varios órdenes de magnitud desde $1,5 \times 10^{-3}$ mut/nt/rep en el fago de ssRNA Q β hasta $1,8 \times 10^{-8}$ mut/nt/rep en el virus de dsDNA Herpes simplex virus tipo 1 (HSV-1) (Drake,

1993; Drake and Hwang, 2005). Este rango corresponde aproximadamente a la fidelidad de las polimerasas usadas en la replicación. Los virus de RNA que utilizan RNA polimerasas dependientes de RNA (RdRps) mutan más rápido que los virus que emplean RNA polimerasas dependientes de DNA (RdDps) o reversotranscriptasas (RTs), que a su vez mutan más rápido que los virus de DNA (con DNA polimerasas). Mediante ensayos *in vitro* se ha confirmado que las RdRps y las RdDps tienen mayor tendencia a error que las DNA polimerasas (Flint et al., 2009). Estas últimas pueden tener dominios de corrección de errores que reducen las tasas de mutación durante la replicación del DNA al menos en un orden de magnitud (Garcia-Diaz and Bebenek, 2007). Sin embargo, no conocemos RNA polimerasas que sean capaces de corregir errores a excepción de las RdRps de miembros de las familias Coronaviridae y Roniviridae que son las únicas que poseen actividad de fidelidad de copia (Minskaia et al., 2006). Por último, la fidelidad de las RTs es mayor que la de las RdRps (aunque sigue siendo inferior que la de las DNA polimerasas), lo que resulta en menores tasas de mutación en los retrovirus que en los virus de RNA.

En la mayoría de los estudios sobre la variación de virus de plantas, la frecuencia de mutación se mide sobre un determinado periodo de tiempo, debido a la dificultad para relacionar la frecuencia de mutación con el número de rondas de replicación del genoma en la planta. A pesar de esta limitación, las frecuencias de mutación pueden usarse para estimar el nivel máximo de la tasa de mutación de un virus de plantas (Sanjuán et al., 2009). Existen dos estudios que han calculado la tasa de mutación de genomas de virus de RNA, uno con TMV (Malpica et al., 2002) y otro con *Tobacco etch virus* (TEV) (Tomas and Elena, 2010). Ambos estudios revelaron una tasa de mutación en torno a 10^{-6} mutaciones por sitio y por generación.

En el caso de los virus de DNA se pensaba que sus bajas tasas de mutación eran debidas a que los replican polimerasas de DNA celulares (Sanjuán et al., 2010). Sin embargo, en virus animales de ssDNA como el parvovirus autónomo carnívoro o el virus adenoasociado de primate se ha observado un alto grado de heterogeneidad genética (López-Bueno et al., 2006). Asimismo, en virus de plantas se han observado niveles elevados de diversidad genética como en los geminivirus *Cotton leaf curl virus* (CLCuV) (Sanz et al., 2000), *Maize streak virus* (MSV) (Isnard et al., 1998), *Tomato yellow leaf curl*

China virus (TYLCCNV) (Ge et al., 2007), y el nanovirus *Faba bean necrotic stunt virus* (FBNSV) (Grigoras et al., 2010), presentando estructuras poblacionales compatibles con el concepto de cuasiespecies similares a las observadas para virus de RNA. Inicialmente, estos elevados niveles de diversidad genética que se detectaban en un huésped se atribuyeron sólo a la frecuente recombinación (Harrison and Robinson, 1999; Hino and Miyata, 2007). Sin embargo, se ha demostrado que los virus de ssDNA parecen tener tasas de sustitución muy cercanas a las de los virus de RNA de tamaño similar, si se comparan con las de los virus de dsDNA como los bacteriofagos de dsDNA, los papilomavirus y poliomavirus, de mayor tamaño (10^{-9} mut/nt/año). Dichas tasas de sustitución de los virus ssDNA se corresponden también con tasas de mutación intermedias entre los virus de RNA y los de dsDNA (revisado en Duffy et al., 2008).

b. Recombinación

La recombinación es el proceso por el cual se obtiene una nueva combinación de una molécula de ácido nucleico. Puede implicar (1) intercambio físico entre cadenas dúplex de secuencias de DNA homólogas, (2) la unión covalente de dos moléculas para generar una nueva a partir de secuencias no homólogas (NHEJ) y la unión de dos secuencias dentro de una misma molécula por medio de repeticiones cortas, en un proceso que ocurre durante la replicación.

Éste es uno de los mecanismos más importantes de generación de nuevos genomas y de adquisición de ventajas selectivas respecto a los genomas parentales. Se ha demostrado que los eventos de recombinación en virus pueden ser la causa de la emergencia de ciertos virus, resultando en un incremento de su gama de huéspedes (Hu et al., 2003; Monci et al., 2002a), o de su virulencia (Rest and Mindell, 2003). El análisis de secuencias de poblaciones de virus de plantas, tanto de RNA como de DNA, indica que la recombinación puede ser una de las mayores fuentes de variabilidad implicadas en su evolución (Chenault and Melcher, 1994; Padidam et al., 1999; Roossinck, 2005).

La recombinación en genomas de RNA suele ocurrir mediante un proceso de selección de copia o molde durante la replicación en el que la polimerasa replicativa salta del molde a una nueva cadena durante la replicación, generando una molécula recombinante entre ambas parentales (Nagy and Simon, 1997; Patil et al., 2007; Hull,

2013). Existen numerosos ejemplos que demuestran la existencia de fenómenos de recombinación en genomas de virus de RNA y DNA de plantas. Así, en virus de RNA pueden mencionarse casos en diferentes géneros: *Luteovirus* (Gibbs and Cooper, 1995), *Potyvirus* (Bousalem et al., 2000; Moreno et al., 2004; Tan et al., 2004), *Tobravirus* (MacFarlane, 1997), *Bromovirus* (Allison et al., 1989), *Closterovirus* (Rubio et al., 2001), *Caulimovirus* (Chenault and Melcher, 1994), *Cucumovirus* (Fernández-Cuartero et al., 1994), *Tobamovirus* (Fraile et al., 1997) o *Potexvirus* (Pagán et al., 2006). A pesar de estos hallazgos, la frecuencia de recombinantes no parece ser muy elevada en las poblaciones naturales de virus de RNA (Roossinck, 2002).

Por el contrario, la recombinación parece ser un fenómeno muy frecuente y de gran importancia en la evolución de virus de la familia *Geminiviridae* (Fauquet et al., 2005; García-Andrés et al., 2006; 2007; Monci et al., 2002a; Richter et al., 2016b). Se han propuesto tres factores para explicar esta alta frecuencia de recombinación: las infecciones mixtas frecuentes, los elevados niveles de replicación y la amplia gama de huéspedes del vector (Padidam et al., 1999). Se dispone de muchos ejemplos en los que se demuestra la existencia de fenómenos de recombinación en geminivirus, tanto entre especies de un mismo género (Bull et al., 2006; Kirthi et al., 2002; Monci et al., 2002a; Ndunguru et al., 2005; Sanz et al., 2000; Saunders et al., 2002), como entre distintos géneros (Briddon et al., 1996; Rojas et al., 2005). La recombinación puede resultar en cambios drásticos de las propiedades biológicas del virus, que pueden dar lugar a un incremento de su ventaja adaptativa (Fernández-Cuartero et al., 1994). Este fenómeno a veces conlleva consecuencias epidemiológicas graves, como la emergencia de nuevos aislados capaces de superar resistencias genéticas o la ampliación de la gama de huéspedes del virus (García-Andrés et al., 2006; García-Arenal and McDonald, 2007; Monci et al., 2002a; Ndunguru et al., 2005). De hecho, varias epidemias de begomovirus se han relacionado directamente con la aparición de virus recombinantes: la enfermedad del mosaico común de la yuca en África (Bull et al., 2006; Pita et al., 2001), la enfermedad del rizado amarillo del tomate (Tomato yellow leaf curl disease, TYLCD) en España, África y la cuenca del Mediterráneo (García-Andrés et al., 2007; Idris and Brown, 2005) y la enfermedad del rizado del algodón en la región Pakistán-India (Idris and Brown, 2005; Lisha et al., 2003; Zhou et al., 1998).

c. Intercambio de componentes genómicos

En los virus de RNA o de DNA con genoma multipartito también puede ocurrir adicionalmente un fenómeno llamado intercambio o reorganización de componentes genómicos, o pseudorrecombinación. Entre los begomovirus (en su mayoría, de genoma bipartito), este proceso puede contribuir a su variabilidad (Hou and Gilbertson, 1996).

d. Segmentación

Recientemente se ha descrito una nueva fuente de variación genética en virus de RNA dependiente de recombinación: la segmentación. El modelo evolutivo que la contempla se basa en la existencia de formas genómicas con deleciones dentro de la cuasiespecie que son infecciosas por complementación (Andino and Domingo, 2015; Domingo et al., 2012). La segmentación ocurre a partir de la selección de genomas cortos cuya replicación se acaba en un menor tiempo que el genoma completo. Un mecanismo plausible para que esto suceda es que las deleciones ocurran de forma aleatoria creando una complementación entre dos zonas del genoma, de forma que se genere una recombinación y que, como resultado de esta, el genoma recombinante supere en número al genoma completo (Domingo et al., 2012).

Existen varios estudios que apoyan la existencia de la segmentación, pero todos ellos han sido realizado *in vitro*. De esta forma estudios con FDMV demostraron la ventaja adaptativa de estos virus con respecto a sus parentales monopartitos (Ojosnegros et al., 2011; Moreno et al., 2014).

OBJETIVOS

El objetivo general de esta tesis es entender la estructura poblacional *in vivo* de las cuasiespecies de virus de RNA y ssDNA cuando se altera su variabilidad genética debido a cambios en la tasa de error, bien por mutagénesis bien por la ausencia de polimerasas mutadoras celulares.

Este objetivo general se ha abordado a través de tres objetivos parciales:

1. Examinar las bases moleculares de la mutagénesis letal *in vivo* de un virus de RNA empleando como modelo el virus del mosaico del tabaco (TMV) en plantas de *Nicotiana tabacum*.
2. Desarrollar un flujo de trabajo bioinformático robusto, versátil y fácil de usar para un análisis integral y pormenorizado de los espectros de mutantes de las cuasiespecies víricas.
3. Analizar la variabilidad genética de virus de ssDNA *in vivo* en ausencia de determinadas polimerasas celulares mutadoras, a saber, las polimerasas de translesión Rev1, PolK, PolH, y la primasa-polimerasa PrimPol en infecciones del geminivirus *Tomato yellow leaf curl virus* (TYLCV) en el hospedador experimental *Arabidopsis thaliana*.

CAPÍTULO 1

Bases moleculares de la mutagénesis letal *in vivo* del virus del mosaico del tabaco (TMV) en plantas de *Nicotiana tabacum* en presencia de 5-Fluorouracilo.

Introducción

1. Catástrofe de error, defección letal y mutagénesis letal como estrategia antiviral

Tal como se indicó en la Introducción General, la composición del espectro de mutantes tiene una gran influencia en el comportamiento de la cuasiespecie como conjunto. La dinámica de las cuasiespecies posee un papel determinante tanto en la patogénesis de los virus como en la prevención y tratamiento de las enfermedades que causan. Una de las implicaciones biológicas de las cuasiespecies víricas es que los espectros de mutantes son reservorios de variantes que en muchos casos pueden emerger como consecuencia de tratamientos antivirales o de la presión del sistema inmune del hospedador. En ocasiones, esas variantes resistentes ya estaban presentes en los espectros de mutantes y se seleccionaron durante los tratamientos antivirales (Domingo et al., 2012; Perales et al., 2012). Una de las estrategias más usadas para el control de las enfermedades virales en plantas es el empleo de cultivares con resistencia genética a determinadas cepas de virus. Sin embargo, dada la naturaleza dinámica de las poblaciones virales, tarde o temprano acaban surgiendo mutantes que son capaces de superar las resistencias de los cultivos causando brotes epidémicos.

Una nueva estrategia antiviral que puede evitar la selección de mutantes resistentes es la mutagénesis letal, fundamentada en la eliminación de los virus RNA por mutagénesis incrementada. Esta terapia antiviral, también conocida como entrada de virus en catástrofe de error, se basa en un concepto teórico, el de umbral de error, introducido originariamente en la teoría de evolución molecular anteriormente mencionada (Eigen, 1971; Eigen and Schuster, 1979). De acuerdo con esta teoría, existe un límite superior en la tasa de error durante la replicación de un organismo que si se traspasase, la información genética se perdería (Eigen, 2002). En el caso de los virus, si se cruzase dicho umbral de error debido a un exceso de mutaciones, la estructura en cuasiespecie no podría mantenerse y se eliminaría la infección vírica. Se ha propuesto otro modelo teórico alternativo para la mutagénesis letal que añade un componente ecológico basado en la tasa de reproducción del virus. No obstante, ambos modelos coinciden en que el virus acabará extinguiéndose por exceso de mutaciones (Bull et al., 2006).

Los primeros experimentos que dieron la idea de que la mutagénesis incrementada podía conseguir la pérdida de infectividad de virus vinieron de la mano de Holland y

colaboradores (Holland et al., 1990) con poliovirus y el virus de la estomatitis vesicular (VSV) empleando análogos de base o nucleósidos mutagénicos. Posteriormente Loeb, Mullins y colaboradores acuñaron el término mutagénesis letal para describir la pérdida de infectividad del virus de la inmunodeficiencia humana adquirida (HIV) en cultivos celulares en presencia de análogo de pirimidina 5-hidroxidesoxicitidina (Loeb et al., 1999). Desde entonces se han documentado pérdidas de infectividad debidas a mutagénesis incrementada, tanto en estudios teóricos como experimentales, en varios sistemas virales animales como poliovirus (Crotty et al., 2000), el virus de la fiebre aftosa (FMDV) (Airaksinen et al., 2003; Pariente et al., 2003; Sierra et al., 2000), el virus de la coriomeningitis linfocítica (LCMV) (Grande-Perez et al., 2002), el HIV (Dapp et al., 2009; Rawson et al., 2016), el virus de la hepatitis C (HCV) (Arias et al., 2014; Contreras et al., 2002; Anderson et al., 2004; Domingo et al., 2006; Ortega-Prieto et al., 2013) entre otros.

Resultados previos de dinámica de cuasiespecies demostraron que el espectro de mutantes puede influir en el comportamiento de cualquier variante de la cuasiespecie viral llegando a suprimirla (La Torre and Holland, 1990). El RNA preextinción obtenido de poblaciones virales sometidas a mutagénesis interfería y retardaba la infección viral influyendo decisivamente en el resultado de la misma (González- López et al., 2004). De acuerdo con el efecto supresor del espectro de mutantes observado para LCMV (Teng et al., 1996), se descubrió en infecciones persistentes en cultivos celulares, que la transición a la extinción debida al tratamiento con el análogo de base mutagénico 5-fluorouracilo (5-FU) estaba caracterizada por la disminución paulatina de la infectividad específica del virus sin una concomitante reducción del número de moléculas RNA genómicas del virus (Arias et al., 2013; Grande-Pérez et al., 2005a; 2005b; Martín et al., 2010; Moreno et al., 2012). El análisis de las moléculas genómicas de RNA de LCMV presentaban mayor complejidad genética pero bajo grado de mutación (Grande-Pérez et al., 2005a) y la secuencia consenso de la población mutagenizada permanecía invariante. Estos resultados dieron pie al desarrollo de un modelo de mutagénesis letal denominado modelo de “defección letal”. Este modelo, obtenido tanto por aproximaciones *in silico* como experimentales en el sistema de LCMV, contempla dos vías a la extinción. En la primera una baja actividad mutagénica enriquece al espectro de mutantes en genomas virales mutados capaces de replicarse, denominados

“defectores”, que mediante una acción interferente conducen a la pérdida de infectividad y al conjunto de la cuasiespecie a su extinción (**Figura 1**). La segunda vía sería bajo una actividad mutagénica muy intensiva que afectaría a la infectividad y a la replicación que se perderían simultáneamente sin la participación de los defectores. El modelo de defección letal ha sido respaldado por varios estudios en cultivos celulares (González- López et al., 2004) y computacionales (Iranzo and Manrubia, 2009).

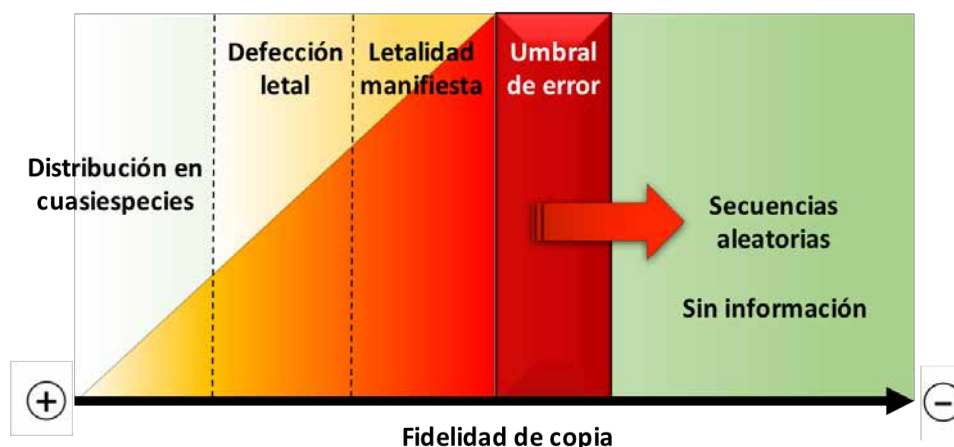


Figura 1. Representación esquemática de los eventos que conducen a la extinción de virus por mutagénesis incrementada. La imagen se basa en el concepto de umbral de error de la teoría de cuasiespecies. El umbral de error se representa aquí como una región donde la fidelidad de copia es tan baja que se produce la transición de una distribución en cuasiespecies a una distribución de secuencias aleatorias carentes de información. Antes de entrar en el umbral de error, suceden al menos dos efectos como consecuencia de la disminución de la fidelidad de copia: defección letal seguida de letalidad manifiesta de la cuasiespecie. La interferencia producida por moléculas defectoras y la presencia de mutaciones letales que conducen a las poblaciones virales hacia la extinción puede lograrse mediante el uso de análogos de base como el 5-FU. Imagen modificada de Domingo *et al.*, 2012.

Otra consideración sobre la dinámica de cuasiespecies virales con grandes implicaciones en patogénesis y en la diseminación viral es la complementación entre componentes del espectro de mutantes. Resultados con poliovirus han demostrado que la patogénesis se determina *in vivo* por la acción cooperativa del espectro de mutantes de la cuasiespecie (Vignuzzi et al., 2006) implicando al conjunto de la cuasiespecie como unidad de selección (Martín et al., 2006; Perales et al., 2005). El comportamiento de los virus RNA se ve influido por interacciones entre genomas víricos y los productos de su expresión dentro del espectro de mutantes de una cuasiespecie vírica que se está replicando. En este sentido, en apoyo del modelo de defección letal para la extinción vírica por mutagénesis incrementada se ha observado que mutantes específicos de la cápside y de

la polimerasa del virus de la fiebre aftosa (VFA) interfieren de manera sinérgica con la replicación del RNA de tipo silvestre (Grande-Pérez et al., 2005a; Perales et al., 2007)., lo que sugiere que el comportamiento de virus altamente variables puede estar fuertemente influido por la composición del espectro de mutantes como un conjunto. Además, se ha demostrado que mutantes dominantes de poliovirus pueden inhibir *trans* la replicación vírica (Crowder and Kirkegaard, 2005). En plantas, el cucumovirus *Tomato aspermy virus* mantiene en su espectro de mutantes un mutante letal por complementación, lo que indica que los virus de plantas son capaces de mantener dentro de un hospedador variantes de bajo *fitness* (Moreno et al., 1997).

A pesar de que la mutagénesis letal se ha estudiado en profundidad en cultivos celulares existen pocos estudios *in vivo* de mutagénesis letal. Algunos compuestos antivirales con actividad mutagénica en cultivos celulares parecen tener la misma actividad *in vivo* aunque el mecanismo terapéutico se desconoce. Por ejemplo la ribavirina, un análogo de purina sintético que es, sin duda, el compuesto antiviral con actividad mutagénica más estudiado (Brochot et al., 2007; Crotty et al., 2000; Maag et al., 2001; Moreno et al., 2011; Perales et al., 2009). La ribavirina se emplea en combinación con interferón para el tratamiento de la hepatitis C aunque no está claro si el modo de acción de la ribavirina sobre HCV *in vivo* se debe a su actividad mutagénica, inmunomoduladora o a otro mecanismo antiviral (Chevaliez et al., 2007; Chung et al., 2013; Dietz et al., 2013; Graci and Cameron, 2005; Lutchman et al., 2007; Perelson and Layden, 2007). Se han documentado aumentos de la frecuencia de mutación de HCV (Asahina et al., 2005; Dietz et al., 2013), aunque otros estudios no han asociado la actividad antiviral de la ribavirina con un aumento de la tasa de error del virus (Chevaliez et al., 2007; Lutchman et al., 2007). Otro análogo mutagénico, el 5-FU, fue capaz de prevenir *in vivo* el establecimiento de una infección persistente del LCMV de ratón (Ruiz-Jarabo et al., 2003). Por otro lado, después de los resultados obtenidos por (Loeb et al., 1999) de mutagénesis letal de HIV-1 en cultivos celulares, el análogo 5,6-dihidro-5-aza-2'-deoxicitidina (KP-1461) llegó a la Fase II de un ensayo clínico en humanos. Si bien dicho análogo alteró el espectro de mutantes de HIV no consiguió un descenso de la carga vírica (Hicks et al., 2013; Mullins et al., 2011). Está por comprobar si el efecto antiviral de KP-1461 observado en cultivos podría ser potenciado *in vivo* en terapia de combinación con otros agentes antivirales. La combinación de dos análogos de citosina,

gemcitabina (2'-desoxi-2',2'-difluorocitidina), un inhibidor de la ribonucleótido reductasa, y decitabina (5-aza-2'-desoxicitidina), un análogo de citosina mutagénico similar a KP-1461, disminuyó la infectividad de HIV un 73% en cultivos y representa una nueva aproximación muy prometedora para la extinción de HIV-1 (Clouser et al., 2010). Dicha terapia de combinación podría ser más efectiva si el tratamiento fuera secuencial, primero el inhibidor (no mutagénico) y a continuación el mutágeno tal como se ha visto para FMDV en cultivos celulares (Perales et al., 2009). En un modelo de ratón ambos análogos en combinación inhibieron la replicación y la progresión de la enfermedad causada por otro retrovirus, el virus de la leucemia murina (MuLV), aunque no se estudió si la actividad antiretroviral iba asociada a un aumento de la carga mutacional en el virus (Clouser et al., 2011). Recientemente se ha investigado la actividad antiviral *in vivo* de otro análogo de base con capacidad mutagénica, el favipiravir (5-Fluoro-2-oxo-1H-pirazina-3-carboxamida) (Arias et al., 2014; Yamada et al., 2016). Este análogo aumentó el número de mutaciones en el RNA vírico y redujo la carga de norovirus en ratones produciendo en algunos casos la extinción del virus.

Lo que se deduce de estos trabajos es que se necesitan más estudios *in vivo* para entender las bases moleculares de la mutagénesis letal para que esta estrategia antiviral se convierta en una terapia real en el tratamiento de las infecciones de virus RNA. Sin embargo, la experimentación *in vivo* con sistemas de virus animales presenta una gran complejidad, un coste muy elevado y el requerimiento de una mano de obra muy especializada. Además de estas limitaciones la experimentación con animales posee una carga ética añadida que dificulta grandemente su realización. En esta Tesis Doctoral hemos elegido un sistema experimental de virus de plantas para estudiar las bases moleculares de la mutagénesis letal. La experimentación con plantas permite llevar a cabo estudios de los aspectos básicos de virus cuyos hospedadores son organismos pluricelulares eucariotas ya que se trabaja con el organismo completo sin grandes impedimentos ni implicaciones éticas. Además, se ha demostrado que los virus de plantas presentan una estructura en cuasiespecies similar al de los virus animales (revisado en Domingo et al., 2012), lo que apoya el uso de estos sistemas modelo para abordar los distintos aspectos de la dinámica de cuasiespecies virales.

2. TMV como modelo de estudio de cuasiespecies virales y mutagénesis letal

El estudio de la enfermedad del mosaico del tabaco (Tobacco Mosaic Disease, TMD), causado por el que posteriormente sería llamado virus del mosaico del tabaco (*Tobacco mosaic virus*, TMV) está en el origen de la Virología como una rama de la ciencia. A finales del siglo XIX, Adolf Mayer, director del *Agricultural Research Station* de Wageningen, investigó dicha enfermedad y en 1886 mostró que podía transferirse inoculando extracto de plantas enfermas a plantas de *N. tabacum* sanas (Mayer, 1886). De esta forma pudo establecer la naturaleza infecciosa de la enfermedad pero no pudo identificar el agente que la causaba. Unos años más tarde, en 1892, Dmitri Ivanowski demostró que los extractos de plantas con TMD pasados a través de filtros capaces de retener bacterias, seguían siendo infecciosos (Ivanowski, 1968) y estableció que la causa de esta enfermedad era un “agente filtrable”. Sin embargo, Ivanowsky concluyó erróneamente que el agente causal era una toxina o un agente bacteriano capaz de pasar a través del filtro Pasteur-Chamberland. Fue Martinus Beijerinck de la *Technical School* de Delft, el que repitiendo y ampliando esos estudios de filtrado dio un salto conceptual en 1898 (Beijerinck, 1898) demostrando que el TMD era causado por un “*contagium vivum fluidum*” o fluido vivo contagioso. Así, TMV pasó a ser el primero de una nueva clase de “agentes filtrables” causantes de enfermedades, cuya nomenclatura fue simplificada al término virus.

Durante la primera mitad del siglo XX, TMV jugó un papel fundamental en el desarrollo de la virología moderna. Esto se debió en parte a la facilidad para infectar y propagar el virus en plantas. Así, TMV fue el primer virus que se purificó en estado cristalino y el primero para el que se demostró una composición basada en ribonucleoproteínas (F C Bawden, 1936; Stanley, 1935). También fue el primero para el que se determinó la arquitectura de sus componentes por medio de la difracción de rayos X y del microscopio electrónico (Bernal and Fankuchen, 1941; Caspar, 1956; Franklin and Klug, 1956). Además, con él se consiguió por primera vez ensamblar *in vitro* el RNA viral y las proteínas de la cápside (Fraenkel-Conrat and Williams, 1955) y fue el primer virus para el que se pudo demostrar que el virión era infeccioso (Gierer and Schramm, 1956).

Desde entonces, TMV ha estado en la vanguardia de la investigación y en el desarrollo de nuevas técnicas y conceptos en la virología de plantas moderna. Por ejemplo, el

desarrollo de técnicas para la obtención de protoplastos para posteriormente infectarlos con TMV (Aoki and Takebe, 1969); el descubrimiento del movimiento célula a célula de las partículas virales de TMV mediante un proceso activo facilitado por proteínas codificadas por el mismo virus (Nishiguchi and Motoyoshi, 1978); la expresión de la proteína de la cápside en la planta, que constituyó el primer ejemplo confirmado de resistencia a patógenos de plantas transgénicas resistentes a TMV (Abel et al., 1986); y muchos de los conocimientos actuales acerca de la función de las proteínas del movimiento (MP) en la alteración del tamaño de los plasmodesmos (Wolf et al., 1989).

3. TMV y el género tobamovirus

TMV pertenece al género *Tobamovirus* junto con otros 5 géneros dentro de la familia *Virgaviridae*, siendo el único de esta familia que no tiene su genoma segmentado. El genoma es de RNA de cadena sencilla (ssRNA) de polaridad positiva, sin segmentar y con un tamaño de unos 6,4 kb (**Figura 2**). Contiene 4 ORFs, que codifican (de 5' a 3'): una metiltransferasa/helicasa, una polimerasa de RNA dependiente de RNA (RdRp) (denominadas como 126K y 183K, respectivamente), la proteína MP (30K) y la proteína

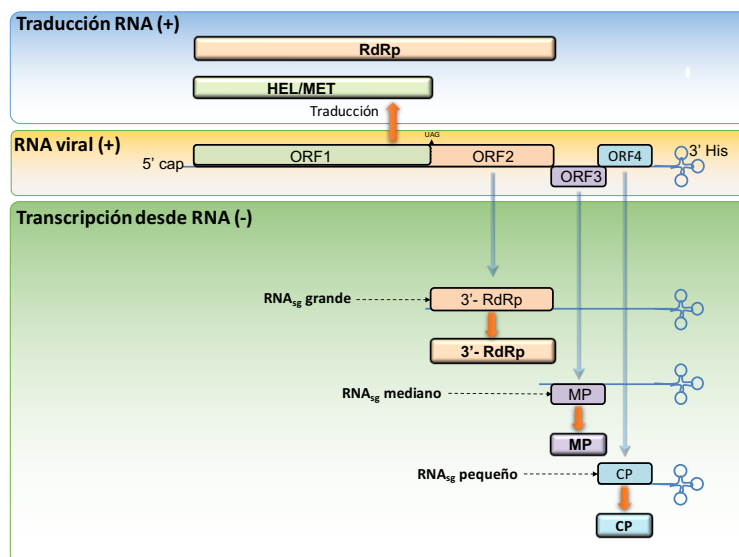


Figura 2. Esquema de la expresión de las proteínas de TMV. El RNA vírico, de polaridad positiva (sobre fondo amarillo) tiene 4 ORFs solapantes que codifican para la subunidad de la polimerasa HEL/MET (ORF1), subunidad grande de la RdRp (ORF2), MP (ORF3) y CP (ORF4). El RNA viral puede actuar como RNAm desde el cual traducir las dos subunidades de la RdRp (fondo azul). Sin embargo, la expresión del resto de proteínas depende de la replicación del RNA (+) a RNA (-) y de la generación de RNAs subgenómicos (RNAsg) por transcripción de la cadena negativa (fondo verde) ya que a partir de estos RNAsg se transcriben las proteínas RdRp, CP y MP.

de la cápside (CP) o 17,5K. El RNA viral de TMV tiene capacidad infectiva (Gierer and Schramm, 1956), gracias a que su genoma funciona inicialmente como RNAm para la traducción de la RdRp (proteínas 126K y 183K).

3.1. Estructura y función de las regiones no codificantes de TMV

El genoma de TMV contiene dos regiones UTR en sus extremos (5'-UTR y 3'-UTR) que están implicadas en el aumento de eficiencia de los procesos de traducción y transcripción (Day et al., 1993; Gallie et al., 1987).

- **Región 5'-UTR**

El extremo 5' terminal está formado por una caperuza de 7-metil guanosina (m⁷GpppG) seguido de una región 5'-UTR rica en AU de unos 70 nts. En muchas cepas de TMV la región 5'-UTR contiene dos zonas muy conservadas formadas por tres copias de ocho nucleótidos que se repiten y una región que contiene repeticiones de la secuencia CAA (Tanguay and Gallie, 1996).

- **Región 3'-UTR**

El extremo 3'-UTR forma una estructura terciaria plegada con cinco pseudonudos, dos de los cuales (los dos últimos) forman una estructura similar a un RNA transferente (RNAt_{his}) (Lewandowski and Dawson, 1999) (**Figura 3**). La estructura RNAt_{his} y el anudamiento más cercano parecen tener un papel importante en la síntesis de RNA de cadena negativa. El extremo 5'-UTR parece interactuar con una región de 72 nts del

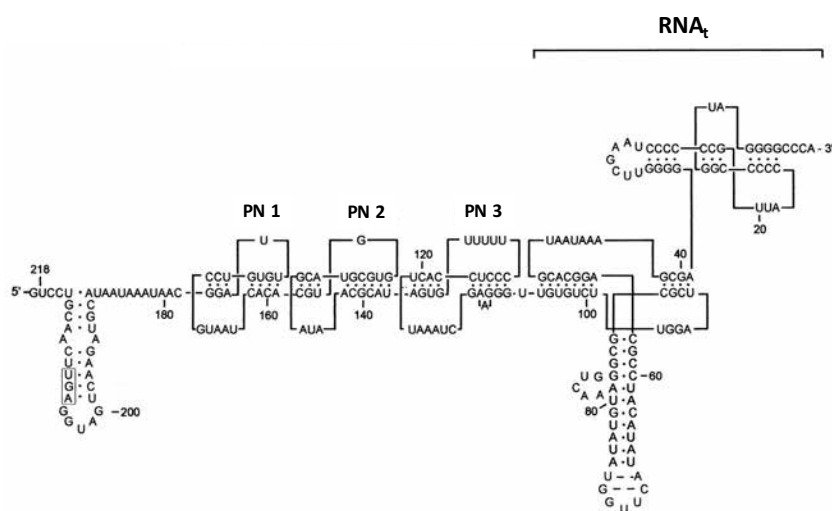


Figura 3. Estructura secundaria del extremo 3'-UTR de TMV. Pueden apreciarse los 3 pseudonudos (PN1,PN2 y PN3) y la conformación RNA_t. Los nucleótidos se han numerado de forma ascendente desde el extremo 3'.

3'-UTR que forma parte de los anudamientos que no forman el RNAt_{his}. Esta interacción es necesaria para una traducción eficiente (Tanguay and Gallie, 1996).

3.2. Estructura y función de las proteínas virales

- HEL/MET (126K)

La proteína 126K (y el extremo amino terminal de la proteína 183 K) contienen dominios conservados típicos de metiltransferasas y helicasas. Se ha demostrado que la proteína 126K es capaz de unir GTP y que mutaciones deletéreas en MP y CP no inhiben la replicación, sugiriendo que 126K junto con (posiblemente) 183K son necesarias para la replicación viral y que estas dos proteínas son subunidades de una polimerasa. Ensayos *in vitro* sugieren que la presencia de 126K parece incrementar la eficiencia de replicación hasta 10 veces.

- RdRp (183K)

La proteína 183K es aparentemente una RdRp cuyo extremo carboxilo terminal contiene un motivo GDD típico de este tipo de polimerasas virales. Puede actuar en *trans* porque los mutantes que expresan solo esta proteína pueden formar todas las formas replicativas del RNA tanto positivas como negativas.

- MP (30 K)

Es una proteína no estructural que se ha relacionado con el movimiento célula a célula y a larga distancia de TMV. MP tiene capacidad de asociarse con los plasmodesmos y estudios *in vitro* indican que es capaz de unirse a ácidos nucleicos de cadena sencilla

- CP (17,5K)

Es la única proteína responsable de formar la envuelta de la partícula vírica. Un virión completo contiene un total de 2130 monómeros de CP. Es responsable del movimiento a larga distancia del virus dentro de la planta infectada. Su estructura secundaria está formada por seis α -hélices.

3.3. Estructura del virión y ciclo de vida

El virión tiene una morfología cilíndrica, alargada y rígida, con aproximadamente 18 nm de diámetro con una cavidad central y con simetría helicoidal. Los viriones pueden tener una longitud variable pero el tamaño predominante está entre los 300-310 nm y dentro de éste se encuentra el RNA genómico (**Figura 4**).

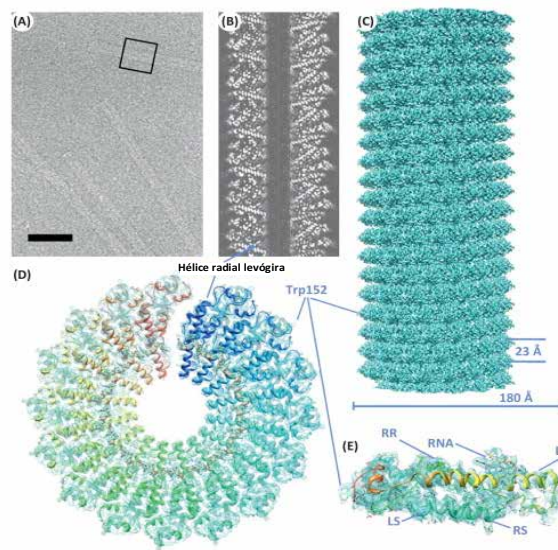


Figura 4. Reconstrucción de la estructura de TMV por crio-microscopía electrónica a 33 nm de resolución. A) Micrografía de viriones de TMV en hielo vítreo (la escala de la barra es de 60 nm). B) Capa central del mapa de la reconstrucción tridimensional de TMV. C) Visión general del mapa de densidad de TMV. D) Giro de la cápside compuesto de 16 subunidades de CP con las α -hélices en arco iris y el RNA unido a las subunidades representado como modelo atómico en el radio interior de la hélice. E) Representación de una CP unida a tres bases del RNA viral. LR, RR, LS y RS: α -hélices de la CP. Adaptado de Ge and Zhou, 2011.

Una vez dentro de la célula vegetal, el comienzo de la desencapsidación viene precedido por la eliminación de algunas subunidades de la CP que envuelven al virión y que contienen el extremo 5' terminal del genoma de TMV (Mundry et al., 1991; Wilson, 1984). Después de este proceso, se desensamblan tres cuartas partes del virión dejando el genoma disponible para que los ribosomas traduzcan de forma simultánea las proteínas 126K y 183K (Shaw et al., 1986; Wilson, 1984; Wu et al., 1994). A continuación se produce otro desensamblaje, esta vez en dirección 3' a 5' asociado a la síntesis de la cadena complementaria de sentido negativo (Wu and Shaw, 1997). Después del desensamblaje inicial, las proteínas 126K y 183K son traducidas directamente desde el RNA genómico. La proteína 183K se produce a partir de la supresión del codón de terminación UAG de la proteína 126K que ocurre entre un 5 y un 10% de los casos (Beier et al., 1984; Pelham, 1978). Como resultado de esta supresión del codón de terminación, ambas proteínas se leen en el mismo marco de lectura y la secuencia de aminoácidos de la proteína 126K está incluida en el extremo amino terminal de la proteína 183K. La eficiencia del salto de codón regula la expresión relativa de ambas proteínas.

Siguiendo a la traducción inicial de estas dos proteínas ocurre la replicación completa de la cadena negativa complementaria de RNA que sirve de molde para la síntesis de copias adicionales del RNA genómico. El dímero de proteínas 126K y 183K reconoce en ellas promotores subgenómicos y se sintetizan hasta tres RNAs subgenómicos (RNAsg), cada uno de los cuales es funcionalmente monocistrónico (**Figura 2**) (Lewandowski and Dawson, 1999; Shaw et al., 1986). El más pequeño de los RNAsg es el más abundante y de él se traduce la CP. El RNAsg de tamaño intermedio se acumula en mucha menor cantidad que el RNAsg de CP y se traduce para dar la proteína MP. Estudios computacionales así como de complementación usando virus mutantes y plantas transgénicas que expresan la MP mostraron que es la única proteína viral esencial para el movimiento de célula a célula (Deom et al., 1987; Meshi et al., 1987). La CP sin embargo, no es requerida para el movimiento de célula a célula pero sí para una infección sistémica de la planta a través del floema (Dawson and Boyd, 1987). Los RNAsg sirven de elementos reguladores de la expresión de la MP y la CP de forma temporal y cuantitativa. Mientras que CP se produce en los últimos estadios de la infección y en cantidades extremadamente altas, MP se sintetiza desde las primeras fases y en baja cantidad, decreciendo su expresión en estadios tardíos de la infección (Atkins et al., 1991; Dawson and Lehto, 1990).

El RNAsg de mayor tamaño encontrado en células infectadas con TMV codifica para una proteína teórica de 54 KDa cuya secuencia sería la misma que el dominio carboxilo terminal de la proteína 183K, aunque esta proteína nunca se ha hallado en células infectadas (Carr et al., 1992; Sulzinski et al., 1985).

Todos estos eventos replicativos de TMV ocurren asociados a la membrana del retículo endoplasmático (RE) perinuclear (**Figura 5**), donde los distintos elementos del citoesqueleto interaccionan con el complejo RNA/RdRp. Los RNA virales actúan como RNAm traduciendo en primer lugar la MP, esta permanece unida con el RNA viral al complejo RNA/RdRp, dando lugar a la formación de estructuras alargadas derivadas del RE. En este punto, la distribución del RNA viral está determinada por un equilibrio entre la formación y ensanchamiento de las mencionadas estructuras y el movimiento del complejo hacia la región periférica de la célula infectada. Una vez ubicado el complejo en la zona periférica, la MP junto con microfilamentos participan en la formación y ensanchamiento de las estructuras derivadas del RE, mientras que los microtúbulos son

responsables del transporte del complejo a través de los plasmodesmos (Guenoun-Gelbart et al., 2008; Más and Beachy, 1999).

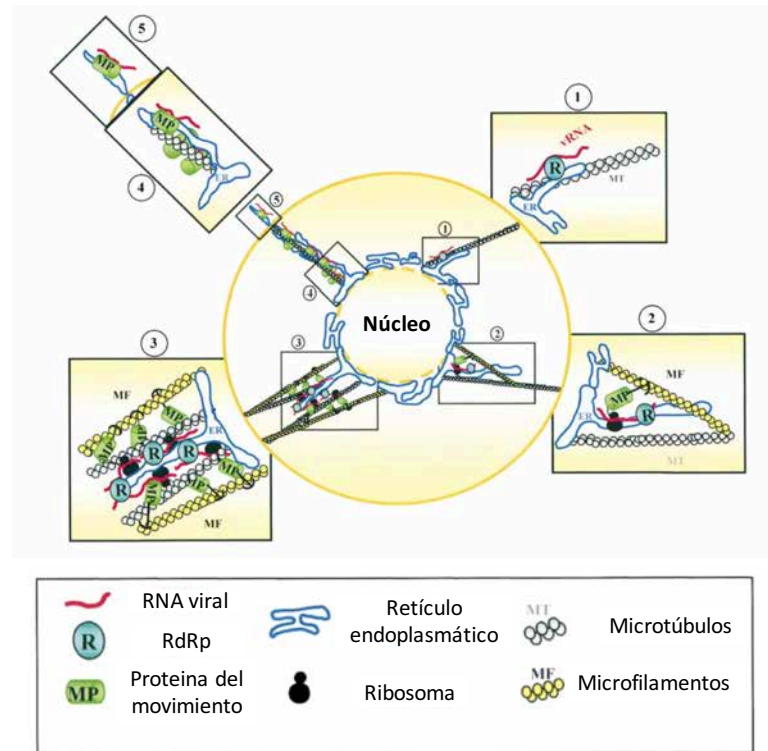


Figura 5. Modelo de infección por TMV en protoplastos de células indiferenciadas BY-2. (1) En la infección temprana, el RNA viral se asocia a la membrana del retículo endoplasmático, y el complejo RdRp/RNAv es transportado a través de microtúbulos a zonas perinucleares, siempre asociado al RE. (2) El RNA viral se replica y se traduce la MP a partir de un RNAsg procedente de la replicación de la cadena negativa que permanece asociada al RE. (3) Las uniones del complejo RdRp/ARNv/MP se estabilizan por medio del anclaje de MP a los microfilamentos asociados al RE. (4) A partir de los microtúbulos se forma un sistema de transporte del complejo RNAv/MP que lleva dicho complejo a la periferia de la célula y pone a éste en contacto con los plasmodesmos para que se inicie la propagación célula a célula. (5) Saliente del RE a la célula adyacente a través de los plasmodesmos con el complejo RNAv/MP asociado. Adaptado de Más and Beachy, 1999.

3.4. Mutagénesis química de TMV

La mutagénesis química aumenta la tasa de error de los virus RNA y se empleó para obtener mutantes de TMV en estudios funcionales de sus proteínas *in vivo*, además de dilucidar muchos otros aspectos importantes de virología de plantas (Hull, 2013). Los estudios de inducción de mutaciones en la proteína de la cápside de TMV han sido de considerable importancia en dos tipos de investigaciones en Biología: en la determinación de la naturaleza del código genético y en la confirmación de la naturaleza

de las mutaciones. Asimismo, se han empleado mutágenos para inducir mutantes sensibles a la temperatura (ts), principalmente en TMV, lo que ha contribuido notablemente a la comprensión de las funciones *in vivo* del genoma viral. Se han empleado también como fuente de cepas de virus de menor virulencia (mild) para control de enfermedades por protección cruzada con dichos mutantes debilitados (Hull, 2013). Gierer y Mundry (Gierer and Mundry, 1958) mostraron que tratamientos de TMV purificado con ácido nitroso *in vitro* incrementaba la frecuencia espontánea de mutantes de lesiones necróticas más de 20 veces.

Otros mutágenos químicos empleados en estos estudios han sido el 5-FU y el 2-tiouracilo (TU), análogos de uracilo que se incorporan al RNA del virus que se está replicando en plantas a las que se les ha suministrado. El 5-FU reemplaza residuos de uracilo en el RNA y puede conducir a los cambios U→C y A→G (Gleason and Fraenkel-Conrat, 1976; Hull, 2013).

Tratamientos con 5-FU *in vitro* redujeron hasta el 90% de la infectividad de TMV (Dawson and Boyd, 1987; Dawson and Lozoya-Saldana, 1984; Dawson and Painter, 1978; Dawson et al., 1976; Holmes, 1955) y presentaron resultados matemáticos y experimentales sugiriendo que en hojas mecánicamente inoculadas la adición de TU prevenía el movimiento célula a célula de TMV. Posteriormente se demostró que este análogo no afecta a la síntesis del RNA viral de TMV sino a alguna otra función (Dawson and Grantham, 1983). En protoplastos de tabaco, (Barta et al., 1981) encontraron que el TU no inhibía ni la incorporación de ³²P en el RNA de TMV ni la multiplicación viral aunque sí disminuía un 10-20% la infectividad específica de TMV extraído de protoplastos tratados con TU, si se comparaba con la de TMV de protoplastos sin tratar. El número de células infectadas en hojas tratadas con TU disminuía en un 60% a un 70% con respecto al hojas no tratadas. Estos autores sugirieron que el TU podía expresar su efecto inhibitorio en las hojas produciendo partículas defectivas de TMV incapaces de infectar célula a célula, sin impedir la inhibición directa de la multiplicación viral. Cabe señalar que estos resultados están en concordancia con nuestro modelo de defeción letal anteriormente expuesto.

Actualmente se desconocen las bases moleculares de la pérdida de infectividad debida a estos análogos en virus de plantas. Para abordarlo hemos seleccionado el 5-FU ya que, como se ha indicado, existen resultados previos que documentan la pérdida de

infectividad de TMV con este análogo de base aunque dicha actividad antiviral nunca se ha asociado a una actividad mutagénica sobre TMV.

3.5. Estructura y metabolismo del 5-FU

El 5-FU es un análogo del uracilo con un átomo de flúor en el carbono 5 en lugar de un hidrógeno (**Figura 6**). Es capaz de entrar rápidamente en la célula utilizando el mismo mecanismo de transporte del uracilo. El 5-FU una vez dentro de la célula es

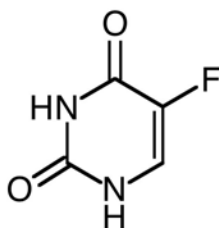


Figura 6. Molécula de 5-FU. En ella puede apreciarse el átomo de flúor en el carbono 5 del anillo aromático.

transformado en tres metabolitos activos: fluorodeoxiuridina monofosfato (FdUMP), fluorodeoxiuridina trifosfato (FdUTP) y fluorouridina trifosfato (FUTP). Estos metabolitos interrumpen la síntesis de RNA y la acción de la timidilato sintasa (TS) (**Figura 7**).

La capacidad del 5-FU de incorporarse al DNA o RNA y unirse a adenina o guanina es debida al átomo de flúor, que retira electrones del anillo provocando, a su vez, que el anillo atraiga electrones de otros átomos, de forma que la molécula adopta la configuración imino (**Figura 8**). En esta configuración, la base no aparea con adenina pero sí con guanina, ya que se forma un tercer puente de hidrógeno debido a la atracción de electrón del flúor. Dependiendo de que configuración (amino o imino) se incorporará una adenina o una guanina a la cadena molde.

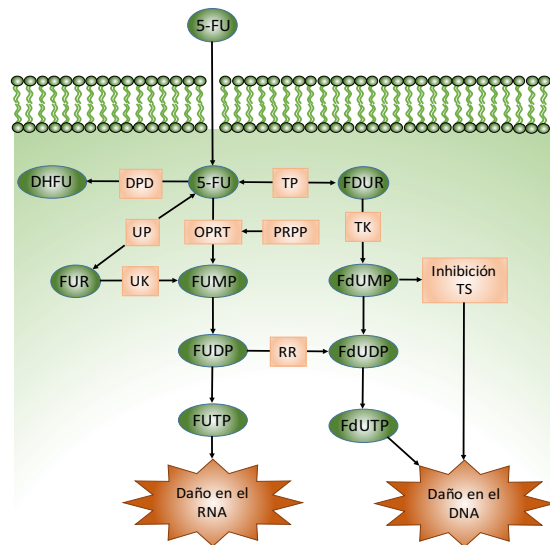


Figura 7. Mecanismo de acción del 5-FU en el interior de las células. El 5-FU es convertido en sus tres principales metabolitos: fluorodeoxiuridina monofosfato (FdUMP), fluorodeoxiuridina trifosfato (FdUTP) y fluorouridina trifosfato (FUTP). El principal mecanismo de activación del 5-FU es su conversión a fluorouridina monofosfato (FUMP), bien directamente a través de la orotato fosforribosil transferasa (OPRT) con fosforribosil pirofosfato (PRPP) como cofactor, o indirectamente a través de fluorouridina (FUR) mediante la acción secuencial de la uridina fosforilasa (UP) y la uridina quinasa (UK). A continuación el FUMP es fosforilado a fluorouridina difosfato (FUDP), que puede ser fosforilado a FUTP, o convertido a fluorodesoxiuridina difosfato (FdUDP) por la ribonucleótido reductasa (RR). A su vez, el FdUDP puede ser fosforilado o desfosforilado para generar los metabolitos activos FdUTP y FdUMP, respectivamente. Una vía alternativa de activación implica la conversión de 5-FU a fluorodesoxiuridina (FdUR) por la timidina fosforilasa (TP), que luego será fosforilado por la timidina quinasa (TK) a FdUMP. Modificado de Longley et al, 2003.

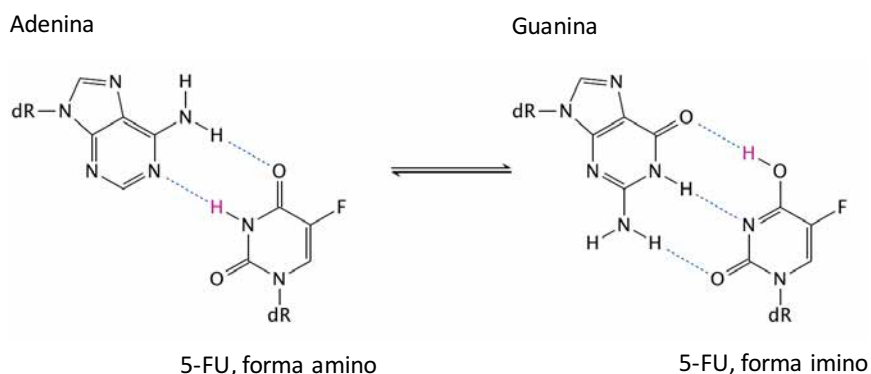


Figura 8. Tautomería de la molécula de 5-FU. En su forma ionizada forma dos puentes de hidrógeno con la adenina y en su forma enol empareja con la guanina mediante 3 puentes de hidrógeno.

- **Inhibición de la timidilato sintasa**

Se ha descrito que el 5-FU es capaz de inhibir la timidilato sintasa (TS) de la célula hospedadora. La TS cataliza la metilación del deoxiuridina monofosfato (dUMP) a desoxitimidina monofosfato (dTMP), utilizado como donador del grupo metilo el metileno tetrahidofolato (CH_2THF). Esta reacción proporciona una fuente de timidilato necesario para la replicación y reparación del DNA. La TS funciona como un dímero, en el que ambas subunidades contienen un sitio de unión a nucleótido y otro sitio de unión a CH_2THF . El 5-FU se metaboliza a FdUMP y se une al sitio de unión del nucleótido en la TS, formando una unión estable con el CH_2THF y bloqueando la unión del sustrato natural de la reacción (dUMP) e inhibiendo, por tanto, la síntesis de dTMP.

Como consecuencia del descenso en la síntesis de dTMP también descienden los niveles de dTTP, lo que termina generando alteraciones en el resto de desoxirribonucleótidos trifosfato (dATP, dGTP y dCTP) debido a los mecanismos que regulan la concentración de éstos dentro de la célula. Estos desbalances en el *pool* de dNTPs son responsables de interrupciones en la replicación y reparación del DNA, resultando en un daño letal para las células a altas concentraciones. La inhibición de la TS lleva consigo una acumulación de dUMP, lo que conduce a un incremento en los niveles de desoxiuridina trifosfato (dUTP), pudiendo incorporarse junto al FdUTP al DNA. La reparación de estas incorporaciones de uracilo y 5-FU a las cadenas de DNA las lleva a cabo la uracil-DNA-glicosilasa (UDG). Sin embargo, esta enzima en presencia de un ratio alto de (F)dUTP/dTTP no es capaz de realizar de forma correcta su función, dando como resultado una “falsa reparación” que termina derivando en roturas de cadena y muerte celular.

- **Incorporación al metabolismo de RNA**

El metabolito del 5-FU FUTP puede incorporarse al RNA, impidiendo el procesado y función de estos. Existen estudios que correlacionan directamente la incorporación de FUTP en el RNA y la pérdida del potencial clonogénico celular. La incorporación de 5-FU puede ser altamente tóxico para el RNA en muchos aspectos, no solo inhibe el proceso de maduración de los pre-RNA a RNAr, sino que también interrumpe las modificaciones postranscripcionales de los RNAt así como el ensamblaje y la actividad de los complejos RNA pequeños nucleares/proteínas, inhibiendo por tanto el *splicing* alternativo del RNAm. Además, los RNAr, RNAt y RNAsn (RNA pequeño nuclear) contienen una base

modificada llamada pseudouridina, siendo el 5-FU capaz de inhibir la conversión postranscripcional de la uridina a pseudouridina en estos RNA. También es responsable de la inhibición de la poliadenilación de los RNAm. Todos estos estudios *in vitro* indican que el 5-FU puede interrumpir casi cualquier aspecto del procesamiento y maduración de los RNAs, ocasionando con ello profundos cambios en el metabolismo celular y en la viabilidad celular.

3.6. Sistema modelo TMV-*N. tabacum* para el estudio de la mutagénesis letal

En esta Tesis Doctoral hemos elegido el sistema experimental de infecciones de TMV en plantas de *N. tabacum* para estudiar *in vivo* los eventos moleculares que conducen a la extinción del virus por la mutagénesis letal. Este sistema nos pareció adecuado por varios motivos. En primer lugar, la heterogeneidad genética, compatible con una estructura en cuasiespecies, de TMV ha sido documentada (Rodríguez-Cerezo and García-Arenal, 1989) así como la tasa de mutación espontánea y espectro mutacional de TMV en este hospedador (Malpica et al., 2002). También se conoce el nivel de diversidad de las cuasiespecies de TMV en *Nicotiana* a partir de un clon infeccioso (Schneider and Roossinck, 2000). Además, se ha estudiado el efecto de los cuellos de botella sobre la variación natural de TMV tras sucesivos pases seriados en el mismo hospedador (Sacristán et al., 2003) u otros hospedadores (Schneider and Roossinck, 2001). Por otro lado, existen plantas de *Nicotiana tabacum* que permiten realizar tanto infecciones sistémicas como infecciones locales. TMV puede infectar sistémicamente plantas de *N. tabacum* cv Samsun nn. Sin embargo, las plantas *N. tabacum* cv Samsun NN, debido a que portan el gen de resistencia N (obtenido de *N. glutinosa*), presentan una respuesta de hipersensibilidad traducida en lesiones locales en el lugar de inoculación. Esto permite emplear este hospedador para obtener clones biológicos y cuantificar virus infecciosos de manera similar a los ensayos en placa para virus líticos animales o bacteriófagos.

En este capítulo hemos estudiado las bases moleculares de la mutagénesis letal de TMV con el análogo de base 5-FU durante el establecimiento de infecciones sistémicas en plantas de *N. tabacum* Samsun nn. Los resultados muestran que el 5-FU disminuyó la infectividad específica de TMV sin afectar a su carga viral. El efecto antiviral no se asoció con aumentos en la frecuencia de mutación ni con la cantidad de mutaciones por

molécula. Sin embargo en las cuasiespecies que mostraron el efecto antiviral del 5- FU hubo diferencias en la complejidad de los espectros de mutantes, el tipo de cambios y la distribución de las mutaciones con respecto al control, aunque todo ello dependió de la región genómica analizada, la RdRp o la MP. No obstante en ambas regiones predominaron las transiciones de base esperadas por la acción del análogo. En conjunto nuestros resultados apoyan el modelo de defección letal *in vivo* del 5-FU sobre TMV y sugieren que el efecto antiviral va asociado a una alteración del equilibrio mutación-selección especialmente en la región RdRp del virus que está más restringida al cambio.

Materiales y métodos

1. Material biológico

1.1. Plantas

Para la realización de este trabajo se han utilizado las plantas *Nicotiana tabacum* cv Samsun nn y *Nicotiana tabacum* cv Samsun NN descritas en la **Tabla 1**.

Tabla 1. Líneas de *Nicotiana tabacum*

Cultivar	Genotipo	Observaciones
Samsun	nn	Plantas que permiten una infección sistémica del virus TMV a partir de la inoculación mecánica. Estas plantas fueron utilizadas para los ensayos <i>in vitro</i> con análogos de base y para la obtención de extractos.
Samsun	NN	Plantas homocigotas para el gen N de <i>Nicotiana glutinosa</i> (Valleau, 1949) que en presencia de TMV desencadena una respuesta de hipersensibilidad y evita la infección sistémica, provocando la necrosis de la zona de la hoja infectada, lo que permite el recuento de lesiones como si de placas de lisis se tratase (Figura 1). Esta línea se utilizó para los ensayos de lesiones locales.



Figura 1. Ensayo de lesiones locales en *Nicotiana tabacum* cv. Samsun NN. Se muestran las lesiones debidas a la respuesta de hipersensibilidad frente a TMV. Las lesiones necróticas aparecen en la media hoja inoculada como placas circulares de color dorado, correspondiendo cada una de ellas a una unidad infectiva de virus.

1.2. Cultivos celulares

Para el análisis de los niveles intracelulares de ribonucleótidos trifosfato (NTPs) y de fluorouracilo trifosfato (FUTP) se utilizaron cultivos celulares de *Nicotiana tabacum* Bright Yellow-2 (BY-2) (DSMZ PC-1181) crecidos a partir de callos BY-2 como los que se muestran en la **Figura 2**. Estas células BY-2 se caracterizan por un rápido crecimiento, de hasta 100 veces la concentración inicial en una semana en condiciones de cultivo óptimas, y son utilizadas comúnmente como modelo en estudios celulares de plantas.



Figura 2. Callo de células *Nicotiana tabacum* Bright Yellow-2. Callo de 30 días, mostrando el fenotipo clásico de esta línea.

1.3. Cepas bacterianas

Para el desarrollo experimental de este capítulo se utilizó la cepa DH5 α de *Escherichia coli* cuya descripción puede encontrarse en la **Tabla 2**.

Tabla 2: Cepas bacterianas.

Especie	Cepa	Genotipo	Observaciones
<i>Escherichia coli</i>	DH5 α	<i>fhuA2 lac(Δ)U169 phoA glnV44 Φ80' lacZ(Δ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17</i>	Cepa receptora utilizada para la clonación de fragmentos de ADN y en la propagación y almacenamiento de plásmidos.

1.4. Stock de TMV

El stock de TMV utilizado en este capítulo procede originalmente del clon infectivo p843pe35TMVr.1 (amablemente cedido por la Dra. Annan Kadri) cuyo mapa mostrando la disposición de los elementos más importantes se muestra en la **Figura 3**.

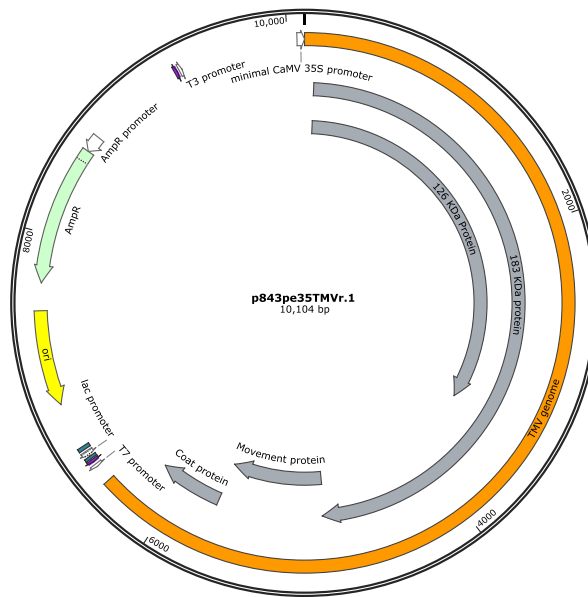


Figura 3. Mapa del clon infeccioso p843pe35TMVr.1. La barra naranja representa la posición del genoma de TMV con los ORFs representados como flechas grises. En verde el gen de resistencia a Ampicilina (Amp^R) y en amarillo el origen de replicación (ori). También se han representado en blanco los promotores existentes en la secuencia de p843pe35TMVr.1

Este clon infeccioso se inoculó en el hospedador de lesiones locales *N. tabacum* cv Samsun NN. El virus de una placa necrótica se inoculó en una planta de *N. tabacum* cv Samsun nn a partir de la cual se obtuvo el stock de TMV con el que se realizó el trabajo de este capítulo. Para ello 20 gr de hojas se congelaron en nitrógeno líquido y se trituraron con mortero con 60 ml de buffer fosfato (10 mM Na₂HPO₄; pH 7,2) que se centrifugó a 10000 g durante 15 min a 4°C. El sobrenadante se fraccionó en alícuotas y almacenó a -80°C hasta su uso.

1.5. Extractos de TMV

Los extractos de TMV para el ensayo de lesiones locales se obtuvieron a partir de plantas *N. tabacum* cv. Samsun nn infectadas. Las plantas se trituraron en nitrógeno líquido y por cada 100 mg de tejido se añadió buffer fosfato en una proporción 1:1 (p/v), mezclando bien y centrifugando a 4°C a 14000 g durante 15 min. La mitad del sobrenadante se empleó para determinar su infectividad por ensayo de lesiones locales y la otra mitad para extracción de RNA.

2. Medios de cultivo

2.1. Medios de cultivo bacteriano

Los medios de cultivo se esterilizaron en autoclave (120°C, 20 min) o mediante filtración. La adición de antibióticos se realizó después del autoclavado y cuando el cultivo alcanzó temperaturas inferiores a los 60°C, en condiciones de esterilidad.

- Medio Luria-Bertani (LB)

Para el crecimiento de *E. coli* o como base de otros medios de cultivo.

Líquido: bacto triptona (10 g/l), extracto de levadura (5 g/l) y NaCl (5 g/l).

Sólido: se añade bactoagar (15 g/l) a la mezcla anterior previamente disuelta.

- Medio X-Gal (LB+IPTG+X-Gal+Amp)

Medio selectivo estándar para experimentos de transformación de la estirpe de *E. coli* DH5 α . Cuando se utilizan vectores con dianas de inserción localizadas en la región codificante del péptido de la β -galactosidasa, la presencia de isopropil- β -D-1-tiogalactopiranosido (IPTG) y X-Gal permite la identificación directa de clones recombinantes por inactivación del gen que codifica dicha enzima por inserción. LB sólido previamente esterilizado y suplementado con 80 μ M IPTG, 0.005 % X-Gal (p/v en formamida), 100 μ g/ml ampicilina en condiciones de esterilidad.

- Medio PSI-b

Para la preparación de células competentes.

Líquido: bacto triptona (20 g/l), extracto levadura (5 g/l), NaCl (0.6 g/l), KCl (0.185 g/l); ajustar el pH a 7. Después de esterilizar en autoclave se añade 10 mM MgSO₄.

2.2. Medios para cultivo de *Nicotiana tabacum*

- Medios estériles

Los medios de cultivo se esterilizaron en autoclave (120°C, 20 min) o mediante filtración. La adición de antibióticos se realizó después del autoclavado y cuando el cultivo alcanzó temperaturas inferiores a los 60°C, en condiciones de esterilidad.

Medio MS+sacarosa

Medio nutritivo para el crecimiento de *N. tabacum in vitro*. Compuesto por: Murashige and Skoog Basal Medium (MS) (4.4 g/L), sacarosa (30 g/L), agar de plantas (8 g/L); Ajustar pH a 5.8 (con KOH).

Medio MS-BY

Medio nutritivo para el crecimiento de células BY-2 en suspensión y para el mantenimiento de callos BY-2.

Líquido: Medio MS; 0,2 g/l de KH_2PO_4 ; ácido 2,4-diclorofenoxiacético al 0,02%, mioinositol al 10% y tiamina al 0,1% (10 mg de ácido 2,4-diclorofenoxiacético, 2,5 ml de etanol de 96, 5 g de mioinositol en 40 ml de agua milliQ más 0,05 g de hidrocloreuro de tiamina). La esterilización se llevó a cabo por filtración.

Sólido: se añade bactoagar (9 g/l) a la mezcla anterior previamente disuelta.

- Medios no estériles

Para el crecimiento de *Nicotiana tabacum* en condiciones no estériles se utilizó una mezcla de turba, fibra de coco y litonita (6:3:1) y se crecieron en semilleros de 10 cm de diámetro.

2.3. Análogo de base

Para la preparación de los medios conteniendo el 5-FU se preparó un stock a 10 mg/ml en agua MilliQ que se esterilizó por filtración y se mantuvo a -80°C hasta su uso. Para los experimentos se realizaron diluciones en medio MS del stock hasta obtener concentraciones finales de 25, 50 y 100 $\mu\text{g}/\text{ml}$.

3. Cultivo y crecimiento de organismos

3.1. Cultivo y crecimiento de bacterias

Escherichia coli se sembró en placas de Petri con medio LB sólido, o líquido en matraces y/o tubos de ensayo con una capacidad cuatro veces mayor que el volumen de cultivo para permitir una buena aireación del mismo en agitación (175-200 rpm). La temperatura de incubación fue de 37°C .

La conservación de las estirpes bacterianas durante largos períodos se llevó a cabo a -80°C en medio LB líquido con 15% glicerol (v/v). Para periodos cortos, se mantuvieron a 4°C en medio LB sólido.

3.2. Cultivo y crecimiento de plantas

- Cultivo de plantas en macetas

La germinación de *Nicotiana tabacum* cv. Samsun NN se llevó a cabo en cámaras húmedas hechas con placas de Petri de 15 cm de diámetro con papel secante dentro previamente humedecido con 5 ml de agua MilliQ estéril, para conseguir una humedad relativa del 100% en su interior, donde se depositaron un número aproximado de 50

semillas. Para sincronizar la germinación de las semillas, las placas húmedas de germinación se incubaron a 4°C durante 48 horas en condiciones de oscuridad, tapadas con papel de aluminio. Tras este tiempo, se colocaron en cámaras de crecimiento con tubos fluorescentes de 36 W a 24°C/16 horas de luz, 18°C/8 horas de oscuridad, en ciclo de día largo. Transcurrida una semana, se trasplantaron a macetas de 15 cm de diámetro x 8 cm de alto. Las plantas se regaron mediante subirrigación con agua cada dos días, aportándoles una solución nutritiva (8,74 mM NO₃⁻; 0,43 mM NH₄⁺; 0,96 mM P; 2,33 mM Ca; 0,74 mM Mg; 0,93 mM S; microelementos comerciales de la casa Tarssan® Mix: B 0,35 ppm; Cu 0,10 ppm; Fe 3,70 ppm; Mn 0,10 ppm; Mo 0,10 ppm; Zn 0,75 ppm) una vez a la semana. Cuando las plantas fueron destinadas a la obtención de semillas, una vez que se marchitaron las corolas florales más jóvenes, se detuvo el riego y se dejaron secar las plantas.

- **Cultivo *in vitro***

Previo a la siembra *in vitro*, las semillas se esterilizaron sumergiéndolas en una solución de lejía 50% (v/v) y de Tween20 0,1% (v/v), manteniéndolas en agitación durante 15 min. Posteriormente, se realizaron 4 lavados consecutivos de 5 min cada uno con agua destilada estéril

- **Cultivo de células en suspensión**

Se partió inicialmente de un callo comercial de células BY-2, que se dispersó con ayuda de unas pinzas estériles en campana de flujo laminar y se resuspendió en 100 ml de medio MS-BY en un matraz Erlenmeyer de 250 ml. Las células se disgregaron con pipeta y se incubaron durante una semana en agitación (100 g) en ciclo de día largo. Partiendo de esta suspensión se realizó el subcultivo celular con 5 ml del stock inicial en 20 ml de medio MS-BY en matraz de 100 ml manteniéndolo en agitación y ciclo de día largo. Los subcultivos se realizaron semanalmente empleando la misma densidad celular.

3.3. Mantenimiento de los callos de BY-2

Se emplearon placas de MS-BY semisólido. Con pinzas estériles se separaron porciones superficiales del callo, evitando tomar las partes en contacto con el medio y se disgregaron con cuidado sobre la superficie del medio fresco. Las placas con los callos se envolvieron con papel de aluminio para asegurar la oscuridad y se incubaron en cámara de cultivo a 22°C. La renovación del medio de los callos se realizó una vez al mes.

3.4. Ensayo de lesiones locales

Para calcular la infectividad de los extractos procedentes de plantas *N. tabacum* cv Samsun nn infectadas con TMV en ausencia o presencia de 5-FU se utilizaron plantas de *N. tabacum* cv Samsun NN. Para ello se partió de 50 μ l de extracto que se diluyeron en 750 μ l de buffer fosfato y a partir del mismo se realizaron tres diluciones seriadas 1:10. A continuación se usaron 30 μ L de estas diluciones para inocular 4 medias hojas de *N. tabacum* cv Samsun NN con cada una de las diluciones, añadiendo un 10% (v/v) de carborundum. Las medias hojas fueron seleccionadas siguiendo un cuadrado latino, tal como se representa en la **Figura 4**, usando un asa de vidrio estéril con la que se extendió el extracto por toda la hoja. Para cada ensayo se emplearon seis plantas, de entre cinco y ocho semanas, a las que se les cortó el ápice, dejándolas con las cuatro hojas más desarrolladas.

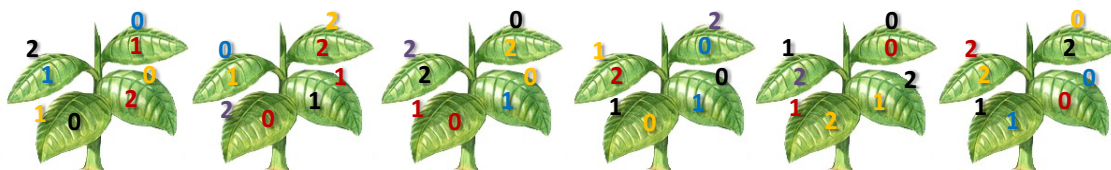


Figura 4. Ensayo de lesiones locales siguiendo el esquema del cuadrado latino. Para ensayar cuatro extractos se emplearon seis plantas. Cada planta (con 4 hojas) es utilizada para realizar 8 inoculaciones, utilizando media hoja para cada una de ellas. Los números representan diluciones del extracto de TMV a ensayar siendo 0 no diluido, 1 la dilución 1:10 y 2 la dilución 1:100. Las diluciones de cada extracto se indican en el mismo color.

Las plantas se mantuvieron en cámaras de cultivo en ciclo de día largo (24°C/16 h luz, 18°C/8 h de oscuridad). Las lesiones se contaron a los 5 días post inoculación (Chapman, 1998).

4. Oligonucleótidos

Los oligonucleótidos utilizados como cebadores en reacciones de amplificación y/o cuantificación, se diseñaron con la ayuda de los programas Primer3 y OligoCalc, analizando en cada caso su estabilidad interna, la formación de dímeros y horquillas no deseados, así como diversos parámetros físicos y químicos (T_m , T_d , % G+C, % A+T). Se obtuvieron de la compañía Sigma Aldrich.

4.1. Cebadores empleados para la retrotranscripción de TMV

Los cebadores utilizados en reacciones de retrotranscripción están indicados en la **Tabla 3**.

Tabla 3. Cebadores usados para síntesis de DNAc.

Cebador	Posición	Secuencia	Descripción
qTMV1699_R	1699-1680	5'-AGCGCAGGCATGTCCACAGA	Utilizado para cuantificación absoluta de TMV extraído de plantas tratadas con el análogo de base 5-FU
A-5684_R	5684-5665	5'-GTAGCCTCCGAATCATCATC	Utilizado para amplificar por PCR Touchdown y clonar el fragmento de TMV 4364-5684.

4.2. Cebadores empleados para RT-qPCR

Los cebadores empleados para la cuantificación absoluta de la carga viral de TMV en plantas de *Nicotiana tabacum* cultivadas *in vitro* en ausencia o presencia de 5-FU están mostrados en la **Tabla 4**.

Tabla 4. Cebadores empleados en la cuantificación absoluta de moléculas de TMV

Cebador	Posición	Secuencia	Descripción
qTMV1699_R qTMVFor1367	1699-1680 1367-1387	5'-AGCGCAGGCATGTCCACAGA 5'-CGGTGTGACAGCGAGGTCCG	Pareja de cebadores que amplifican una región de 332 nts en la región de la RdRp.
25S_Universal-F 25S_Universal-R		5'-ATAACCGCATCAGGTCTCCA 5'-CCGAAGTTACGGATCCATT	Pareja de cebadores empleados para la normalización de la cuantificación absoluta de moléculas de TMV

4.3. Otros cebadores

Además de los cebadores utilizados para la retrotranscripción y para la cuantificación también se utilizaron cebadores para secuenciación y para la obtención de los estándares para la cuantificación. Estos cebadores están descritos en la **Tabla 5**.

Tabla 5: Otros cebadores utilizados durante el desarrollo experimental

Cebador	Posición	Secuencia	Descripción
A-4364_F	4364-4384 (TMV)	5'-CGCAGGTATCAAACTTGCAT	Cebado empleado junto con A-5684_R para amplificar la región 4364-5684 de TMV por PCR touchdown para generar los clones que se secuenciaron con posterioridad.
1225_F	1225-1245 (TMV)	5'-AGAGGACGCGCAAGGAAGTCT	Cebador con unión a la región RdRp de TMV utilizado en la generación de estándares para cuantificación absoluta de TMV.
pGEMT-2923F	2923-2953 (pGEM-T)	5'-GTGCTGCAAGGCGATTAAGT	Cebador usado con la pareja qTMV1699_R para amplificar la región T7 de pGEM-T junto con el inserto de TMV, generando un amplicón de 631nt.
T7	-	5'-TAATACGACTCACTATAGGG	Cebador con unión al promotor T7 de <i>pGEM-T</i> , empleado para la secuenciación de la región 4364-5684 de TMV.
SP6	-	5'-ATTTAGGTGACACTATAG	Cebador con unión al promotor SP6 de <i>pGEM-T</i> , empleado para la secuenciación de la región 4364-5684 de TMV.

5. Extracción de nucleótidos y ácidos nucleicos

5.1. Extracción de nucleótidos para análisis de HPLC

- Extracción a partir de tejido de plantas

Para llevar a cabo la extracción de NTPs intracelulares a partir de tejido vegetal pulverizado en nitrógeno líquido se siguió el protocolo de extracción de nucleótidos descrito en Meyer y Pogolotti, con modificaciones. Se partió de 4 gr de tejido pulverizado en un tubo de 15 ml al que se le añadieron 4 ml de ácido tricloroacético (TCA) 0,6 M y se incubó 10 minutos en hielo. A continuación se centrifugó a 14800 g durante 10 min a 4°C y se pasaron 4 ml del sobrenadante a un tubo nuevo, previamente enfriado. El sobrenadante fue dividido en 4 tubos eppendorf de 1.5 ml. En el siguiente paso se lavaron columnas de Strata Phenil (55 µm, 70 Å) (Phenomenex) con 3 ml de metanol con un flujo de 3 ml/min con ayuda de una jeringuilla y se equilibró añadiendo 2 ml de TCA 0,6 M. A continuación se aplicó a la columna 1 ml de muestra y se eluyó con TCA 0,6 M recogiendo 4 ml del eluido. Posteriormente 4 ml de una mezcla 2:1 (v/v) de Freón:Trioctilamina (TOA) 98% se pasaron por la columna a 1 ml/min, recogiendo todo el eluido en un tubo de polipropileno de 5 ml previamente enfriado. Tras mezclar durante 15 seg con vórtex la muestra se centrifugó a 2500 g durante 2 min a 4°C, y se

transfirió la fase acuosa a un nuevo tubo comprobando que su pH era de 6,5. Por último la muestra se dividió en fracciones de 1 ml que se congelaron en nitrógeno líquido y se liofilizaron por completo en una SAVANT SpeedVac con trampa de refrigeración. Las muestras se resuspendieron en un volumen de 400 µL de agua MilliQ, se filtraron por 0,22 µm a través de filtros DURAPORE® y se guardaron a -80°C hasta su análisis por HPLC.

- **Extracción de nucleótidos de células de *N. tabacum* BY-2**

Para la extracción de nucleótidos de cultivos celulares BY-2 se partió de 1 ml de un subcultivo saturado de células que se lavaron dos veces con PBS a 4°C y al que se añadieron 500 µl de TCA 0,6 M. Tras 10 min de incubación en hielo 500 µl del sobrenadante se pasaron con cuidado a un nuevo tubo al que se añadieron 500 µl de Freón:TOA 4:1 (v/v) y la mezcla se agitó vigorosamente. Posteriormente se centrifugó a 14500 g durante 30 seg a 4°C y se recogieron 350 µl de la fase acuosa que se filtró y conservó a -80°C hasta su análisis por HPLC.

5.2. Extracción de RNA total de *Nicotiana tabacum* cv Samsun nn

Se partió de 150 mg de muestra pulverizada en nitrógeno líquido en un tubo de eppendorf de 1,5 ml. A cada muestra se le añadieron 300 µl de *Tri Reagent Solution* (Ambion), se agitó la muestra durante 2 min y se centrifugó durante a 12300 g durante 5 min a 4°C, pasando el sobrenadante a un nuevo tubo. A este nuevo tubo se le añadieron 60 µl de cloroformo, se agitó durante 15 s y se dejó reposar 3 min a temperatura ambiente. A continuación se centrifugó a 16000 g 15 min a 4°C. El sobrenadante se transfirió a un tubo nuevo al que se le añadieron 150 µl de isopropanol. Después de 10 min de reposo a 4°C se centrifugó a 12220 g durante 10 min a 4°C. El sobrenadante se transfirió a un nuevo tubo al que se añadieron 300 µl de etanol al 70% y se mezcló con vortex 15 s. Las muestras se centrifugaron a 8800 g 5 min a 4°C, se descartó el sobrenadante y se secó el pellet a temperatura ambiente. Finalmente el pellet se resuspendió en un volumen de 25 µl de agua DEPC.

6. Tratamiento enzimático de ácidos nucleicos

6.1. Tratamiento con DNasa

Como parte de la generación de los estándares para la RT-qPCR descritos en el apartado 8.4. de Materiales y Métodos se eliminó el DNA molde mediante tratamiento con DNasa con el kit *Turbo DNA-free*TM (Ambion). Para cada muestra se añadieron 0,1 volúmenes

de tampón 10 *Turbo DNase* y 1 µl de Turbo DNasa (10 u/µl). La mezcla se incubó 30 minutos a 37°C. Posteriormente se añadió el reactivo de inactivación de DNasa y se incubó 5 min a temperatura ambiente, mezclando ocasionalmente. Por último, se centrifugó la mezcla durante 2 min a 11300 g y el sobrenadante fue transferido a un nuevo tubo y almacenado a -80°C hasta su uso.

6.2. Clonaje de fragmentos de DNA a vectores plasmídicos

Para clonar fragmentos de DNA amplificados por PCR se empleó el *pGEM®-T Easy Vector System* (Promega), siguiendo las instrucciones del fabricante. Los productos de PCR se purificaron utilizando el kit *QIAquick Gel Extraction* (Qiagen). Puesto que la DNA polimerasa *Pfu* empleada en la PCR no incorporaba una adenina terminal al fragmento de PCR, antes de realizar la ligación con el vector se realizó un *tailing* del fragmento de DNA amplificado, es decir, la adición de una adenina terminal mediante la DNA polimerasa BioTaq (0,1 U; Biotline) en presencia de dATP (0,2 mM) durante 30 min a 70°C. La unión de fragmentos de DNA a vectores linearizados se llevó a cabo utilizando el kit *Rapid DNA Ligation* (Roche) que contiene la DNA ligasa del fago T4 (Roche), siguiendo las indicaciones del fabricante.

7. Amplificación de ácidos nucleicos

7.1. PCR estándar

Para la amplificación estándar de fragmentos de DNA se empleó el kit Biotaq DNA polimerasa termoestables de BIOLINE siguiendo las condiciones de reacción que se indican en la **Tabla 6**. La reacción se realizó con 1,5 pmol/ml de cada cebador, una concentración 4 mM de MgCl₂, 0,2 mM de dNTP y 1 U de polimerasa. La cantidad de DNA molde empleada fue de 100-200 ng.

Tabla 6. Etapas de la reacción en cadena de la polimerasa (PCR)

Etapas	Temperatura	Tiempo	Nº de ciclos
Desnaturalización inicial	94°C	5 min	1
Desnaturalización	94°C	30 s	30
Unión al cebador	55°-65°C ^a	30 s	
Extensión	72°C	T_{ext}^b	
Extensión final	72°C	10 min	1
Mantenimiento	10°C	indefinido	1

^aLa temperatura es dependiente de la T_m del cebador, por regla general se usó la relación T=T_m-4°C.

^bEl tiempo de extensión de la reacción, directamente proporcional al tamaño amplificado, fue de 30 s por Kb.

7.2. Síntesis de cDNA mediante transcripción inversa

Se utilizó el kit *AMV Reverse Transcriptase* junto con el kit *Recombinant RNasin® Ribonuclease Inhibitor* (ambos de Promega). A una dilución que contenía 2 µg de RNA total se le añadió 1 µg de cebador en un volumen final de 11 µl. A continuación las muestras se calentaron 5 min a 70°C y después se mantuvieron otros 5 min en hielo. Posteriormente se le añadió un mix que contenía 5 µl de buffer de reacción, 2,5 µl de dNTP mix 10 mM, 40 U de RNasin Ribonuclease inhibitor y 30 U de AMV RT. La reacción se llevó a cabo 60 min a 42°C. Las reacciones se guardaron a -80°C.

7.3. PCR Touchdown

Esta variante de la PCR estándar fue empleada para amplificar cDNA para favorecer la hibridación de los cebadores asumiendo que pudiera existir alguna base desapareada. Esto se consigue mediante la reducción de la especificidad de unión de los cebadores disminuyendo gradualmente la temperatura de hibridación a lo largo del proceso de la PCR. Para llevar a cabo la PCR Touchdown se utilizó la DNA polimerasa de alta fidelidad *Pfu* (Promega), siguiendo las recomendaciones del fabricante. Para ello se partió de 2 µl de cDNA al que se le añadieron 48 µl de un mix que contenía 5 µl de Buffer *Pfu* 10x, 1 µl de dNTPs 10 mM, 0,5 µl de polimerasa *Pfu*, 37,5 µl de agua MilliQ y 2 µl de cada cebador (A5684R y A4364F) a una concentración de 10 µM. Estos cebadores amplifican una región de 1320 nt que cubre un fragmento del extremo 3' de la RdRp y el ORF de la proteína del movimiento casi al completo (**Figura 6**). Las condiciones de la reacción de PCR se indican en la **Tabla 7**.

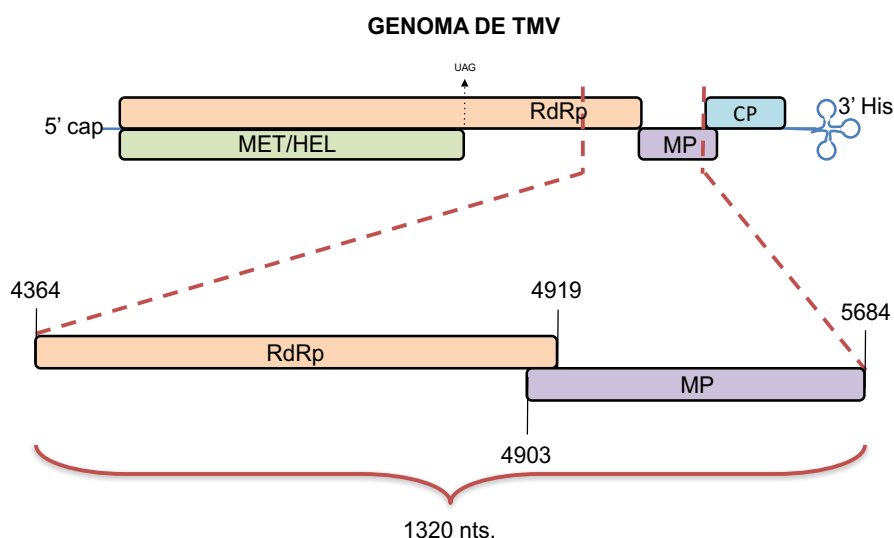


Figura 6. Esquema del genoma de TMV indicando la región analizada. La parte superior de la imagen representa una molécula de ARN genómico de TMV con la región amplificada y secuenciada entre líneas discontinuas rojas. La parte inferior de la figura es una ampliación de la región analizada de TMV donde se muestran con números las posiciones de la región amplificada (del 4364 al 5684) y el inicio y fin de la región solapante de los ORFs RdRp y MP.

Tabla 7. Etapas de la reacción de PCR *Touchdown*

Etapas	Temperatura	Tiempo	Nº de ciclos
Desnaturalización inicial	95°C	2 min	1
Unión al cebador	61°C - 0,6°C/ciclo	30 s	20
Extensión	72°C	2 min	
Desnaturalización	94°C	1 min	
Unión al cebador	48°C	30 s	10
Extensión	72°C	2 min	
Extensión final	72°C	10 min	1
Mantenimiento	10°C	indefinido	1

7.4. RT-PCR cuantitativa (RT-qPCR)

- RNA estándar de TMV para cuantificación absoluta

Para la obtención del estándar de RNA para la recta patrón de la RT-qPCR cuantitativa absoluta de TMV se realizó una amplificación de la región de 474 nts dentro del ORF de la RdRp y del ORF de la subunidad Hel/Met del clon infectivo de TMV p843pe35TMVr.1 utilizando los cebadores 1225_F y qTMV1699_R. Este producto de PCR fue separado por electroforesis en gel de agarosa y el DNA del tamaño esperado se purificó del gel y se clonó en pGEM-T como se ha indicado anteriormente. A continuación se amplificó por PCR el promotor T7 de pGEM-T junto con el inserto con los cebadores pGEMT-2923F y

qTMV1699_R purificando el amplificado nuevamente a partir de gel. La **Figura 7a** muestra el resultado de esta amplificación. Las muestras fueron sometidas a un proceso de transcripción con la RNA polimerasa de T7 para la obtención de RNA de TMV del amplicón de 474 nt como se muestra en la **Figura 7b**. Tras un tratamiento con DNasa fue cuantificado por espectrofotometría en Nanovue® (GE Healthcare) y almacenado a -80°C hasta su uso (**Figura 7c**).

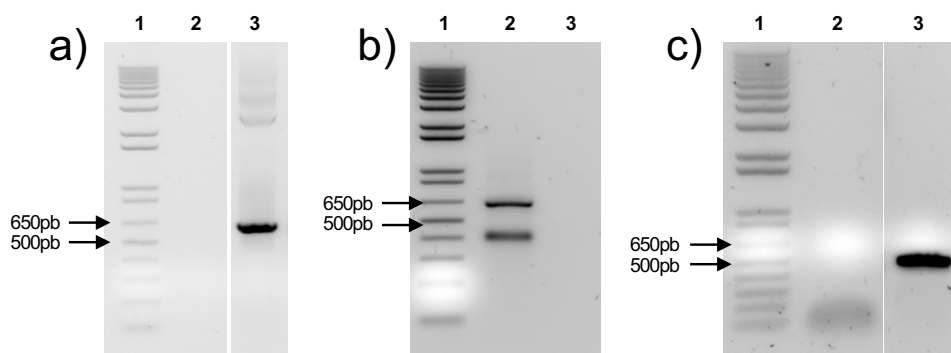


Figura 7. Obtención del estándar de RNA para la cuantificación de TMV. a) Amplificación del clon infectivo de TMV p843pe35TMVr.1 con la pareja de cebadores pGEMT-2923F y qTMV1699_R. Se muestran el marcador 1Kb plus (1), el control negativo (2) y la muestra amplificada (3). b) Producto de la transcripción *in vitro* del producto de DNA amplificado en el paso anterior, en la calle 2 se observan dos bandas correspondientes al DNA molde (alrededor de 650pb) y al RNA transcrito (por debajo de 500pb); 1 Kb Plus (1); control negativo (3). c) RNA tras tratamiento con DNasa. 1 Kb Plus (1), control negativo (2) y muestra tratada con DNasa (3).

Las reacciones de RT-qPCR se llevaron a cabo en dos pasos. En el primero se realizó una RT tal como se indica en el apartado 8.2. El segundo paso se realizó por duplicado (2 repeticiones técnicas a partir de cDNA de cada muestras) en un termociclador LightCycler 480 (Roche), utilizando la mezcla de reacción *Sybr Green I Master* (Roche). El volumen final de reacción fue de 10 µl en lugar de los 20 µl indicados por el fabricante, previa comprobación de su funcionamiento óptimo. Se utilizaron como molde 1 µl de cDNA de cada muestra diluido a 130 ng/µl y se añadieron el resto de componentes de la mezcla de reacción siguiendo las instrucciones del fabricante. Las condiciones de la PCR son las que se detallan en la **Tabla 8**.

Tabla 8. Etapas de la RT-PCR cuantitativa

	Etapas^a	Temperatura	Tiempo	Nº de ciclos
RT	Unión al cebador	70°C	5 min	1
	Extensión	42°C	60 min	1
	Mantenimiento	10°C	Indefinido	1
qPCR	Desnaturalización inicial	95°C	30 s	1
	Desnaturalización	95°C	10 s	
	Unión al cebador y	60°C ^b	15 s	40
	Extensión			
	Curva de disociación	60°C+0.5°C/ciclo	10 s	60

^aLa recogida de datos de fluorescencia se efectúa en las etapas de unión al cebador y en cada ciclo de la curva de disociación (*melting*).

^bTodas las parejas de cebadores fueron diseñados para usar la misma temperatura de unión (*annealing*).

7.5. Cuantificación absoluta

Para la cuantificación absoluta de TMV, se preparó una recta patrón a partir del RNA estándar de TMV descrito en el apartado anterior. Para ello se hicieron diluciones decimales seriadas desde 10^9 a 10^3 copias por μL de RNA calculado con la fórmula:

$$n^{\circ} \text{ de copias de ARN} = \frac{N \cdot 6,022 \times 10^{23}}{T \cdot 1 \times 10^9 \times 330}$$

donde N=ng/ μL y T=pares de base del fragmento de DNA.

La pareja de cebadores qTMVFor1367 y qTMV1699_R se comprobó por análisis de regresión lineal de los valores del ciclo umbral (C_q , *quantification cycle*) de tres réplicas de las diluciones estándar sobre el logaritmo de la cantidad de DNA total de cada muestra. La eficiencia de los cebadores fue del 98% y la R^2 de 0.998.

La normalización de los datos de cuantificación absoluta para cada muestra se llevó a cabo basándose en los valores C_q resultantes de la amplificación por duplicado del gen de referencia correspondiente al codificante para el RNA ribosómico 25S con los cebadores 25S_Universal-F y 25S_Universal-R (**Tabla 4**) siguiendo el protocolo descrito por Mason y colaboradores (Mason et al., 2008).

7.6. Amplificación por círculo rodante (RCA)

Para amplificar los plásmidos de las colonias de *E. coli* transformadas con pGEMT®-TMV(4364-5684) que resultaron positivas (y que mostraron color blanco y resistencia a ampicilina) previa comprobación por PCR, se utilizó la amplificación por círculo rodante (*Rolling circle amplification*, RCA), con el kit *Illustra™ ThempliPhi* (GE Healthcare). Este kit utiliza la DNA polimerasa del bacteriófago $\Phi 29$, cuya característica principal es su alta procesividad y una tasa de error próxima a 1×10^{-7} errores por nucleótido amplificado (Dunning et al., 1988), para amplificar de forma preferente genomas circulares a partir de cebadores con secuencias aleatorias de 6 nucleótidos. Para ello, la colonia bacteriana se tocó con la punta de la pipeta y se mezcló con 5 μ l de *Sample Buffer*. Tras un calentamiento a 95°C durante 3 min, se dejó 5 min en hielo y se añadió una mezcla de 5 μ l de *Reaction Buffer* y 0,2 μ l de *Enzyme mix*. La mezcla se centrifugó 15 s a 1800 g y se incubó a 30°C durante 8 horas. Finalmente la reacción se interrumpió calentando las muestras a 65°C durante 10 min.

8. Electroforesis de ácidos nucleicos

8.1. Electroforesis en geles de agarosa

La separación e identificación de moléculas de DNA se realizó mediante electroforesis horizontal en geles de agarosa (Agarose SPI, Duchefa) disuelta en tampón TBE (Tris-Base 89 mM, EDTA 2 mM y Ácido Bórico 89 mM). La concentración de agarosa varió en función del tamaño de los fragmentos a separar, oscilando entre 0,8-1% (p/v). Para poder visualizar los ácidos nucleicos por transiluminación con luz azul, se añadió Sybr Safe (1 μ l/ml) a la solución de agarosa fundida. Cada muestra se mezcló con tampón de carga para geles de agarosa en proporción 5:1 (v/v). La electroforesis se realizó en cubetas horizontales con tampón TBE, a un voltaje constante de 3-4 V/cm.

Los geles se fotografiaron con un documentador de geles ChemiDoc XRS+ (Biorad) y el programa Image Lab. Como patrones de tamaño molecular se empleó 1 Kb Plus DNA Ladder (Invitrogen).

La concentración de DNA se determinó midiendo la densidad óptica (DO) a una longitud de onda de 260 nm por espectrofotometría o por electroforesis en geles de agarosa en comparación con cantidades conocidas de patrones de DNA.

8.2. Purificación de los fragmentos de DNA de geles de agarosa

Se realizó usando el kit *QIAquick Gel Extraction* (Qiagen), siguiendo las recomendaciones del fabricante. Las soluciones incluidas en el kit proporcionan la concentración salina y pH adecuados para la unión del DNA a una membrana de sílice, y su posterior elución en condiciones básicas.

9. Transformación genética

9.1. Preparación de células competentes

Una colonia fresca de *E. coli* DH5 α se inoculó en 5 ml de PSI-b (ver apartado 2.1. de Materiales y Métodos) y se incubó toda la noche con agitación a 37°C. Todo el cultivo estacionario se transfirió a 100 ml de PSI-b en un matraz de 1 l (precalentado a 37°C) y se continuó la incubación con agitación vigorosa hasta que la DO₆₀₀ fue 0,45-0,5. A continuación el cultivo se enfrió en hielo y se centrifugó a 5000 g durante 5 min a 4°C. Seguidamente, las células se resuspendieron suavemente en 40 ml de solución Tfb-1 (30 mM acetato potásico pH 6,9; 50 mM MnCl₂; 100 mM KCl; 100 mM CaCl₂; 15 % glicerol (v/v)) fría. Tras una incubación de 5 min en hielo se volvió a centrifugar a 5000 g durante 5 min a 4°C, se retiró el sobrenadante y las células se resuspendieron suavemente en 4 ml de solución Tfb-2 (10 mM MOPS pH 7; 75 mM CaCl₂; 10 mM KCl; 15 % glicerol (v/v).) fría. Por último, esta suspensión celular se incubó en hielo durante 15 min, se distribuyó en fracciones alícuotas de 150 μ l, se congeló rápidamente en N₂ líquido y se conservó a -80°C hasta su uso.

9.2. Clonaje molecular para el análisis genético de poblaciones de TMV

A 50 μ l de células *E. coli* DH5 α competentes se añadieron 5 μ l de ligación y la mezcla se incubó 30 min en hielo. Transcurrido ese tiempo, las células se sometieron a un choque térmico de 2 min a 42°C e inmediatamente se volvieron a dejar en hielo durante 5 min. Por último, las células se incubaron a 37°C con agitación durante 30 min en 1 ml de LB. Tras centrifugar, las células se resuspendieron en 150 μ l de medio LB. Por último toda la mezcla de transformación se sembró en placas de LB+IPTG+X-Gal+Kan y se incubaron a 37°C.

9.3. Infección de *N. tabacum*

Plantas de *Nicotiana tabacum* cv Samsun nn crecidas en tarros Magenta® hasta el estadio de 3-4 hojas se inocularon con TMV a una MOI (multiplicidad de infección) de

aproximadamente 50 unidades formadoras de lesión (lfu) por planta un día después del tratamiento con análogo de base 5-FU. Para ello se preparó el inóculo diluyendo el virus stock ($2,54 \times 10^6$ lfu/ml) en tampón fosfato (10 mM NaH_2PO_4 pH 7,2). Sobre la hoja más desarrollada y de fácil acceso de la planta *in vitro* se extendieron 5 μl de la solución infectiva y se distribuyeron por toda la hoja con ayuda de dos asas de siembra de vidrio. Para su posterior identificación se realizó un pequeño agujero en la hoja. Todo este proceso se realizó en condiciones de esterilidad. El mismo día de la inoculación se realizó comprobación del título del stock de virus empleado mediante ensayo de lesiones locales.

10. Análisis de niveles de nucleótidos intracelulares

Las muestras con los nucleótidos intracelulares extraídos según el apartado 6.1. de Materiales y Métodos (100 μl) se inyectaron en un sistema HPLC Alliance 2695 conectado a un detector de matriz de fotodiodos 2996 (Waters). Los nucleótidos se separaron en una columna analítica Partisil 10X SAX (4.6x250 mm) (Whatman) con una precolumna Partisil 10X SAX (4.6 x 30mm) (Capital HPLC). La tasa de flujo de elución fue de 0,8 ml/min y los nucleótidos se detectaron a una longitud de onda de 254 nm. Antes de la inyección de las muestras, la columna se equilibró con tampón A (7 mM $\text{NH}_4\text{H}_2\text{PO}_4$, pH 3,8). El programa para la separación comenzó con 22 min de un periodo isocrático (periodo en el que se utiliza un solo solvente) con tampón A, seguido de un gradiente lineal de 112 minutos hasta alcanzar el 100% de tampón B (250 mM $\text{NH}_4\text{H}_2\text{PO}_4$, 500 mM KCl pH 4,5). Finalmente, se acabó con otro periodo isocrático de 37 minutos con tampón B. Como control de separación y cuantificación se inyectaron 50 μl conteniendo 20 pmol/ μl de CTP, ATP, GTP, UTP y 100 pmol/ μl de FUTP. Para el análisis se utilizó el programa *Waters Empower Chromatography Data Software*. Las determinaciones se llevaron a cabo en tres réplicas biológicas independientes con tres réplicas técnicas cada una de ellas. La cantidad de cada nucleótido en los extractos se hizo relativa al total de nucleótidos medidos en cada muestra según tratamiento y tiempo.

11. Secuenciación

Las colonias de *E.coli* transformadas que crecieron en medio de LB+IPTG+X-Gal+Kan como se indica en el apartado 10.2. de Materiales y Métodos se comprobaron mediante

PCR para confirmar que tanto el inserto como el tamaño del mismo eran correctos. Las colonias positivas se amplificaron por RCA, tal como se indicó en el apartado 8.5. de Materiales y Métodos. Cada reacción de RCA de 10 µl fue dividida en dos y se le añadió a cada una un cebador que hibridase en ambos extremos del inserto de TMV, estos fueron los cebadores SP6 y T7 a una concentración de 10 µM, por último se llevó el volumen final de la reacción a 10 µl y se enviaron a secuenciar a la empresa MacroGen por el método de secuenciación automática Sanger.

12. Caracterización de la variabilidad genética de las cuasiespecies.

Para analizar el efecto del 5-FU sobre las diferentes poblaciones virales se estudió la complejidad y heterogeneidad de las cuasiespecies del ensayo mediante el cálculo de la frecuencia de mutación, diversidad nucleotídica y índice de Shannon normalizado.

12.1. Frecuencias de mutación mínima y máxima

La frecuencia de mutación mínima es el número de mutaciones distintas, incluyendo transiciones, transversiones, deleciones e inserciones, dividido por el número de nucleótidos secuenciados. Las mutaciones repetidas se contaron una sola vez, interpretando que todas las mutaciones iguales derivan de la replicación de un único evento de cambio. Los haplotipos son el número de secuencias distintas dentro de una población.

12.2. Diversidad nucleotídica

La diversidad nucleotídica es la media del número de nucleótidos distintos entre pares de secuencias tomadas al azar del mismo espectro de mutantes y de la misma región genómica. Se calculó según la fórmula:

$$\pi = \frac{n(n-1)}{2 \sum_{ij} \pi_{ij} (i < j)}$$

En la que “n” es el número de clones, y “ π_{ij} ” es el número de nucleótidos distintos entre un par de secuencias “ij” dividido por la longitud de la secuencia en nucleótidos (Nei & Li, 1979). Este cálculo se realizó a través de MEGA 6 y los intervalos de confianza para la diversidad nucleotídica se calcularon usando el método estadístico de remuestreo *bootstrap*.

12.3. Índice de Shannon normalizado

El índice de Shannon normalizado (S_n) se utilizó para medir la heterogeneidad del espectro de mutantes, tomando valores comprendidos entre 0 (homogeneidad completa) y 1 (heterogeneidad completa). La justificación teórica de este concepto se describe en (Volkenstein, 1994) y la aplicación al estudio de las cuasiespecies virales puede encontrarse en (Airaksinen et al., 2003; Pariente et al., 2003; Ruiz-Jarabo et al., 2000; Sierra et al., 2007; 2000). La entropía de Shannon normalizada se calculó a partir de la siguiente fórmula

$$S_n = \frac{-[\sum_i (p_i \cdot \ln p_i)]}{\ln N}$$

donde " p_i " es la frecuencia de cada secuencia en el conjunto de secuencias estudiado y " N " es el número total de secuencias analizadas.

13. Estadística

Todos los análisis estadísticos fueron realizados con el software IBM SPSS v. 22.0 (SPSS Inc.). El estudio de diferencias entre pesos de las muestras tratadas (fresco y seco) se llevó a cabo mediante un análisis estadístico ANOVA con un $P < 0,01$. Para comprobar si había diferencias significativas en la infectividad, en la carga viral y en la infectividad específica se llevó a cabo el análisis no paramétrico de Kruskal-Wallis de un factor para muestras que no siguen una distribución normal con un nivel de significación $P < 0,01$. Para analizar la aleatoriedad de la distribución de las mutaciones en función de la posición genómica, se llevó a cabo un test no paramétrico de rachas con un nivel de significación $P < 0,01$. En el caso de las diferencias en las frecuencias de mutación de muestras entre las diferentes regiones en presencia o ausencia de tratamiento, se aplicó un análisis no paramétrico χ^2 con un grado de libertad y con un $\alpha = 0,05$.

14. Bioinformática

Para el procesado y manejo de los datos obtenidos en el transcurso del trabajo, así como para el análisis de las secuencias estudiadas, se han utilizado distintos programas informáticos, reflejados en la Tabla 9.

Tabla 9. Programas de ordenador empleados en este capítulo

Programa	Empresa/investigador	Aplicación
Seqman (Lasergene 7)	DNASTAR	Alineamiento y detección de variantes de secuenciaciones Sanger.
SPSS 22	SPSS Inc.	análisis estadísticos
Microsoft Excel 15.16	Microsoft	Cálculos y análisis estadísticos
Primer3	H. Hughes Medical Institute	Diseño de cebadores
MEGA 6	(Tamura et al., 2013)	Análisis filogenético y cálculo de distancias genéticas
Snappgene 2.3.2	GSL Biotech	Creación de gráficos de plásmidos
Lightcycler software 1.5	ROCHE	Análisis de datos de cuantificación absoluta
APE v2.0.49	Wayne Davis	Visualización de secuencias, análisis de sitios de restricción, diseños de clonación.
ImageLab 4.1	Bio-Rad	Tratamiento de captura de imágenes de geles de agarosa
ARLEQUIN 3.1	(Excoffier et al., 2005)	Cálculo AMOVA

Resultados

1. Toxicidad del análogo de base 5-FU en plantas de *Nicotiana tabacum* cv Samsun

Debido a la escasez de datos acerca del efecto fitotóxico del 5-FU *in vivo*, realizamos un estudio previo para analizar si las dosis empleadas afectaban al desarrollo vegetativo de plantas de *N. tabacum* cv Samsun nn. Para ello, se crecieron plantas de *N. tabacum* *in vitro* en medio MS sin análogo hasta que las plantas tuvieron cuatro hojas verdaderas. Entonces las plantas se trataron sin (Control) o con 25, 50 o 100 µg/ml de 5-FU durante 10 días. Una vez transcurrido ese tiempo se determinó el peso fresco y el peso seco de entre 18 y 23 plantas por tratamiento. Como se puede observar en la **Figura 1**, no se observaron diferencias significativas en peso fresco ni en peso seco entre las plantas sin tratamiento y las tratadas con las distintas concentraciones de 5-FU según un ANOVA de una vía para $p < 0,05$. Los resultados descartan un efecto tóxico del 5-FU *in vivo* al menos durante 10 días de tratamiento a las concentraciones indicadas.

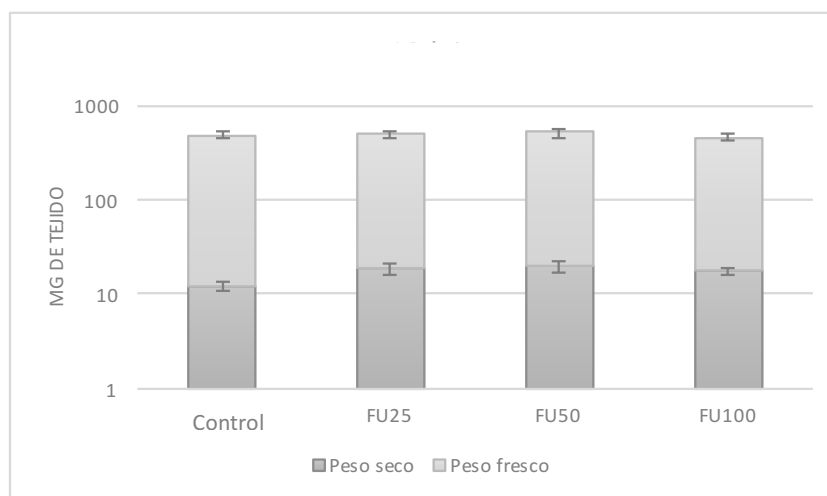


Figura 1. Toxicidad *in vivo* del 5-FU. Se determinó el peso fresco y el peso seco por mg de tejido de plantas *N. tabacum* cv Samsun nn crecidas *in vitro* sin tratar (Control) o tratadas con 25 µg/ml (FU25), 50 µg/ml (FU50) y 100 µg/ml de 5-FU (FU100) durante 10 días. Para cada tratamiento y tiempo se analizaron entre 18 y 23 plantas.

2. Efecto del 5-FU sobre la infectividad y la acumulación de TMV

Se utilizaron plantas de *N. tabacum* cv Samsun nn crecidas *in vitro* durante cuatro semanas en estadio de cuatro hojas verdaderas. Las plantas se crecieron en ausencia

(plantas control) o en presencia de 25, 50 y 100 µg/ml de 5-FU. A las 24 h de comenzar el tratamiento todas las plantas se infectaron con TMV mediante inoculación mecánica del virus stock a una MOI de 50 unidades formadoras de lesión (lfu) por planta. A los 5 y 10 días postinoculación (dpi) se recogieron muestras. Entonces plantas control, así como de cada tratamiento, se trasplantaron a maceteros con sustrato sin análogo. Estas plantas se recogieron a los 31 dpi, después de 21 días *ex vitro*.

2.1. Infectividad

Para comprobar si el análogo 5-FU podía tener un efecto *in vivo* sobre la infectividad de TMV tras los distintos tratamientos de las plantas *N. tabacum* infectadas, realizamos determinaciones de su infectividad. Para ello, se llevaron a cabo ensayos de lesiones locales en hojas de *N. tabacum* cv Samsun NN a partir de extractos obtenidos de las plantas infectadas y cultivadas *in vitro* en ausencia o presencia de 25, 50 y 100 µg/ml de 5-FU a 5, 10 dpi, y de las plantas trasplantadas a maceta tras 10 días de tratamiento *in vitro* y 21 días sin análogo (31 dpi), tal como se indica en el apartado 4 de Materiales y Métodos. En la **Tabla 1** se muestra el número de réplicas biológicas utilizado para cada tiempo y tratamiento.

Tabla 1. Plantas tituladas mediante ensayo de lesiones locales

DPI	CONTROL	FU 25	FU 50	FU 100	Total
5	5	6	11	13	35
10	6	4	4	8	22
31	5	3	3	4	15
Total	16	13	18	25	72

Los resultados de la **Figura 2A** muestran que a 5 dpi no existieron diferencias en el número de unidades lfu/ml entre el control y los distintos tratamientos con el análogo de base. A 10 dpi (**Figura 2B**) se observa una disminución de un 72% con respecto al control en el título vírico de las muestras tratadas con 100 µg/ml de 5-FU, lo que supone una disminución de más de 10^6 partículas de TMV infeccioso. El estudio estadístico no paramétrico de Kruskal-Wallis para comprobar la significación de dichas diferencias confirmó este hecho para las muestras tratadas con 100 µg/ml de 5-FU a 10 dpi ($H=12,025$ con 3 grados de libertad, $P=0,007$). Al eliminar el análogo de base durante 21 días se observó una recuperación del título viral en las muestras previamente tratadas

a 10 dpi con 5-FU, llegando incluso a superar los niveles del control (**Figura 2C**). Sin embargo, las diferencias no fueron significativas según el estudio estadístico de Kruskal-Wallis para $P < 0,05$. Los resultados sugieren un efecto antiviral sobre la infectividad de TMV dependiente de la dosis de 5-FU con 10 días de tratamiento aunque es estadísticamente significativo sólo a la dosis de 100 $\mu\text{g/ml}$ de análogo.

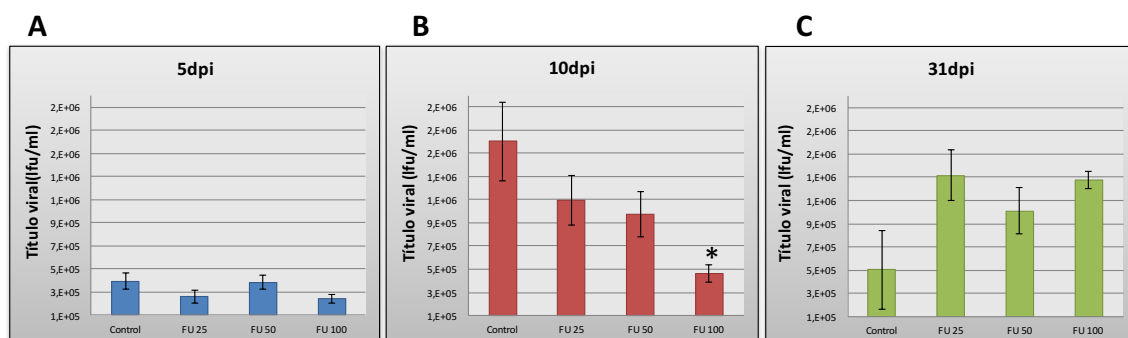


Figura 2. Disminución de la infectividad de TMV en presencia de 5-FU. Se determinó el título viral (Ifu/ml) de las plantas sin tratar (Control) y tratadas con 25, 50 y 100 $\mu\text{g/ml}$ de 5-FU a 5 dpi (a), 10 dpi (b) y 31 dpi (c) mediante ensayo de lesiones locales en plantas de *N. tabacum* cv Samsun NN. La significación estadística (*) se calculó mediante el test Kruskal-Wallis para distribuciones no paramétricas. Las barras de error indican la desviación estándar de las réplicas biológicas.

2.2. Determinación de la carga vírica

Para averiguar si el efecto observado del 5-FU sobre la infectividad de TMV podía deberse a que el análogo podía estar reduciendo la cantidad de moléculas genómicas, y por tanto de viriones infecciosos, determinamos la carga vírica. Para ello se realizaron extracciones de RNA total de los extractos de las plantas infectadas con TMV control y tratadas con 25, 50 y 100 $\mu\text{g/ml}$ de 5-FU a 5, 10 dpi y las trasplantadas a maceta tras 10 días de tratamiento *in vitro* y 21 días sin análogo (31 dpi), tal como se indica en Materiales y Métodos (apartado 6.2). Los RNAs genómicos de TMV se cuantificaron mediante RT-qPCR absoluta en dos pasos. La carga viral alcanzó valores máximos a 5 dpi en todas las muestras analizadas (**Figura 3**), con alrededor de 10^9 moléculas/mg de tejido. Dicha acumulación de genomas de TMV se mantuvo a 10 dpi en todas las muestras y descendió un orden de magnitud a los 31 dpi, en las muestras con un régimen de 10 días de tratamiento seguido de 21 días sin tratamiento.

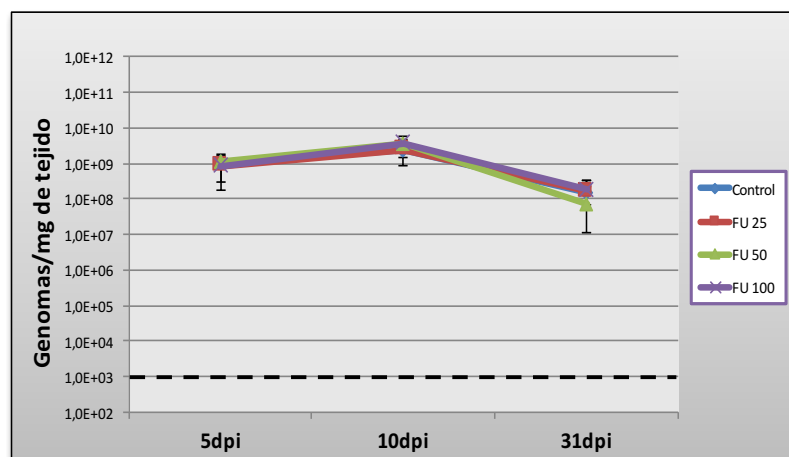


Figura 3. Cuantificación de moléculas genómicas de TMV en plantas *N. tabacum*. Se determinó por RT-PCR cuantitativa absoluta el número de moléculas genómicas de RNA vírico en extractos de plantas *N. tabacum* cv Samsum nn sin tratar (Control) o tratadas con 25 (FU25), 50 (FU50) y 100 (FU100) $\mu\text{g/ml}$ de 5-FU. La línea negra discontinua paralela al eje de abscisas marca el límite de detección de la técnica para cada muestra ($1,9 \times 10^2$ moléculas/mg de tejido). Las barras de error indican la desviación estándar de las réplicas biológicas.

No se observaron diferencias en la carga vírica entre el control y las muestras tratadas con 25, 50 y 100 $\mu\text{g/ml}$ de 5-FU tras 5 o 10 días de tratamiento para un $p < 0,05$. De la misma manera, tampoco se encontraron diferencias en la acumulación de moléculas genómicas de TMV entre control y ninguno de los tratamientos a los 31 dpi ($p < 0,05$). Los resultados muestran que la disminución de la infectividad de TMV en presencia de 5-FU no va acompañada de una reducción de la acumulación de moléculas genómicas, lo que sugiere que la replicación del virus no parece verse afectada por el análogo.

2.3. Determinación de la infectividad específica

El descenso en la infectividad específica, es decir, la infectividad de cada molécula de RNA vírico, es una de las características de la mutagénesis letal mediada por análogos de base o de nucleósido mutagénicos. Dicho descenso se detecta en las cuasiespecies víricas pre-extinción mientras la secuencia consenso del virus se mantiene invariante (González- López et al., 2004; Grande-Pérez et al., 2005a). Para investigar cómo los tratamientos con 5-FU *in vivo* afectaban a la infectividad específica de TMV, realizamos una comparación de la infectividad específica de las cuasiespecies de TMV presentes en las plantas infectadas en ausencia y presencia de diferentes concentraciones de análogo tras 5 y 10 dpi, y tras un régimen de 10 días de tratamiento seguido de 21 días en

ausencia de análogo (31 dpi). Como se muestra en la **Figura 4**, la infectividad específica de TMV descendió en todas las muestras tratadas a 5 y 10 dpi a cualquier dosis comparada con el control, siendo significativo a 100 $\mu\text{g}/\text{ml}$ de 5-FU ($p < 0,005$). Por el contrario, cuando las plantas infectadas tratadas con 5-FU fueron trasplantadas y crecidas en ausencia del análogo de base durante 21 días, se observó un incremento de la infectividad específica en los tres tratamientos que igualó los valores del control. Nótese que la infectividad específica del control fue en aumento durante toda la infección lo que sugiere un proceso de adaptación del virus stock (ver origen el virus stock descrito en el apartado 1.4. de Materiales y Métodos). Tal aumento de infectividad específica se observó también en las poblaciones víricas tratadas con 5-FU una vez retirado el análogo. Los resultados sugieren que la disminución de la infectividad específica se debe a la presencia del 5-FU. En conjunto, los resultados muestran que la acción antiviral del 5-FU sobre TMV *in vivo* presenta características típicas de la mutagénesis letal, es decir, disminución de la infectividad, aunque no de la carga viral, y una bajada de la infectividad específica del virus.

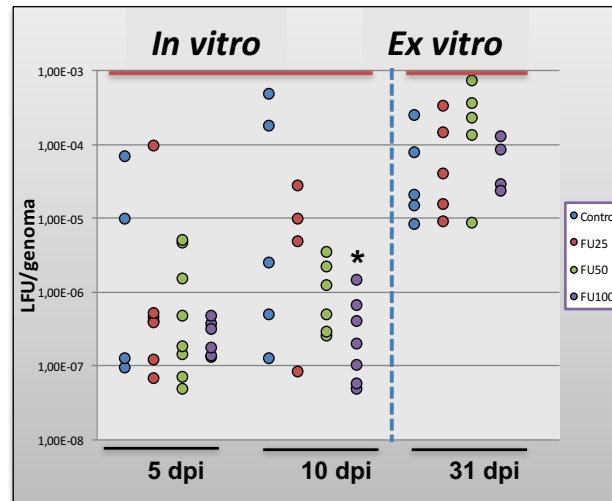


Figura 4. Determinación de la infectividad específica de las poblaciones de TMV sometidas a tratamientos con 5-FU. La infectividad específica (infectividad por molécula de RNA genómica medida en unidades formadoras de lesión, o lfu, por genoma) de cada espectro de mutantes sin tratar (Control) o tratado con 25, 50 y 100 $\mu\text{g}/\text{ml}$ de 5-FU se indica con un punto. La línea discontinua indica el cambio de régimen de cultivo de plantas tratadas 10 días *in vitro*, a *ex vitro* (21 días) en sustrato sin 5-FU. Las diferencias estadísticamente significativas se indican con un asterisco ($p < 0,005$).

3. Caracterización genética de los espectros de mutantes

Para comprender las bases moleculares de la disminución de la infectividad específica de TMV debida al 5-FU, analizamos la variabilidad genética de algunas cuasiespecies de TMV tratadas y sin tratar con este análogo de base. Para ello, partiendo de entre 10^7 y 10^9 moléculas genómicas, se amplificó por RT-PCR en dos pasos una región de 1300 nucleótidos que abarca desde el extremo 3' del ORF de RdRp hasta el extremo 3' del ORF de MP (**Figura 6** de Materiales y Métodos). Las secuencias consenso de las cuasiespecies se obtuvieron mediante secuenciación del DNA amplificado por RT-PCR como se indica en el apartado 8.4. de Materiales y Métodos. Para analizar los espectros de mutantes el DNA amplificado se clonó y aproximadamente 20 clones moleculares de cada muestra se secuenciaron. En total analizamos 15 espectros de mutantes, 345 clones moleculares y secuenciamos un total de 419051 nt (lista de mutaciones de cada cuasiespecie en la **Tabla A1** del Anexo). De las 15 cuasiespecies analizadas sólo se observó un cambio G5626A en la secuencia consenso que correspondió a una planta tratada con 100 µg/ml de 5-FU a 10 dpi (réplica 1). Al analizar el espectro de mutantes observamos que dicho cambio estaba presente en 16 de 20 clones analizados en dicha cuasiespecie (**Tabla A1**). En la **Tabla 2** se muestra el número de mutaciones encontradas en las cuasiespecies control y tratadas con 25, 50 y 100 µg/ml de 5-FU a 5, 10 y 31 dpi. Estos datos nos permitieron estudiar la complejidad así como la heterogeneidad de las cuasiespecies secuenciadas.

Tabla 2. Análisis de clones moleculares de cuasiespecies de TMV control o tratadas con 25, 50 y 100 µg/ml de 5-FU

	Muestra ^a	Clones	Mutaciones	Nucleótidos secuenciados	Frecuencia de mutación ^b	Mutaciones por molécula
5 dpi	Control (1)	25	20	31868	$6,28 \times 10^{-4}$	0,80±0,25
	Control (2)	45	28	57093	$4,90 \times 10^{-4}$	0,55±0,10
	FU25 (1)	25	11	30331	$3,63 \times 10^{-4}$	0,44±0,14
	FU50 (1)	18	8	21178	$3,78 \times 10^{-4}$	0,44±0,14
	FU100 (1)	20	7	21430	$3,27 \times 10^{-4}$	0,35±0,15
	FU100 (2)	28	6	31464	$1,91 \times 10^{-4}$	0,26±0,01
	FU100 (3)	27	12	29987	$4,00 \times 10^{-4}$	0,44±0,12
10 dpi	Control (1)	20	9	25600	$3,52 \times 10^{-4}$	0,65±0,18
	FU25 (1)	18	8	22566	$3,55 \times 10^{-4}$	0,44±0,17
	FU50 (1)	21	11	26277	$4,19 \times 10^{-4}$	0,52±0,16
	FU100 (1)	20	14	24951	$5,61 \times 10^{-4}$	0,75±0,26
	FU100 (2)	20	3	24592	$1,22 \times 10^{-4}$	0,20±0,09
	FU100 (3)	19	6	22641	$2,65 \times 10^{-4}$	0,32±0,13
31 dpi	Control (1)	20	5	24753	$2,02 \times 10^{-4}$	0,25±0,12
	FU100 (1)	19	10	24320	$4,11 \times 10^{-4}$	0,53±0,22
total		345	158	419051		

^aNúmero de la réplica biológica indicado en paréntesis.^bMutaciones por nucleótido secuenciado. Cada mutación se computó una sola vez en cada espectro de mutantes.

4. Complejidad de las cuasiespecies de TMV sometidas a tratamiento con 5-FU

La complejidad de los espectros de mutantes, entendida como la composición mutacional dada por el conjunto de la cuasiespecie, fue analizada mediante el cálculo de la frecuencia de mutación y la distancia genética media. Para el cálculo de la frecuencia de mutación de cada cuasiespecie se cuantificaron las mutaciones con respecto a su consenso y cada mutación se consideró una sola vez. Como se puede apreciar en la **Tabla 2**, la frecuencia de mutación osciló entre $1,91 \times 10^{-4}$ mut/nt para la réplica 2 del tratamiento a 100 µg/ml de 5-FU y $6,98 \times 10^{-4}$ mut/nt para la réplica 1 de la muestra control, ambas a 5 dpi. No se observaron diferencias significativas en la frecuencia de mutación a ninguna concentración de 5-FU, ni siquiera en las plantas que se trataron con 5-FU durante 10 días, seguidas de crecimiento *ex vitro* sin tratar 21 días (Kruskal-Wallis $p < 0,05$). Tampoco hubo diferencias entre las frecuencias de mutación de cada una de las dos regiones genómicas, RdRp y MP (resultados no mostrados). Además,

el número medio de mutaciones por molécula fue inferior a 1 en todas las muestras control y tratadas con 5-FU a cualquier tiempo analizado (**Tabla 2**), lo que indica un bajo grado de mutación por genoma, algo característico de poblaciones víricas en vías de extinción durante la mutagénesis letal.

Como se indicó anteriormente, otro parámetro que refleja la complejidad genética de la cuasiespecie es la distancia genética media (d). Este valor d hace referencia a la media del número de mutaciones existentes entre parejas de secuencias de la población elegidas al azar (Domingo et al., 2006). Para este estudio realizamos alineamientos de las secuencias para cada una de las regiones genómicas RdRp y MP por separado y calculamos d utilizando el método de 2-parametros de Kimura (Kimura, 1980). Estos valores de d pertenecientes a las cuasiespecies de TMV se muestran en la **Tabla 3**.

En la región RdRp tanto a 5 como a 10 dpi los valores medios de d en presencia de 5-FU fueron de 3,3 a 3,96 veces inferiores al control. Por el contrario en la región MP a 5 dpi observamos que los valores medios de d fueron 175 veces menores en las muestras tratadas con 100 $\mu\text{g/ml}$ de 5-FU mientras que a 25 y 50 $\mu\text{g/ml}$ fueron similares al control. A 10 dpi la media de d de las muestras a 25 y 50 $\mu\text{g/ml}$ presentaban 3 y 10 veces, respectivamente, mayor complejidad genética que el control. Sin embargo para 100 $\mu\text{g/ml}$, prácticamente se equiparaba al control. Curiosamente, la réplica 2 del control fue 6,2 y 14,4 veces más diversa que la réplica 1 para las regiones RdRp y MP, respectivamente.

Al realizar una comparación entre regiones genómicas, observamos que la media de la distancia genética de la región MP fue en general superior a la de la RdRp tanto a 5 como a 10 dpi, excepto en el caso de las muestras tratadas con 100 $\mu\text{g/ml}$ de 5-FU a 5 dpi. Estas diferencias entre regiones desaparecen completamente tras 21 días en ausencia de tratamiento. La complejidad en esta muestra de 100 $\mu\text{g/ml}$ fue el doble que en el control.

Tabla 3. Distancia genética media (d) de las regiones RdRp y MP de las cuasiespecies de TMV secuenciadas

	Muestras ^a	RdRp				MP				MP/RdRp ^d
		d	s.e. ^b	Media ^c	SE ^b	d	s.e. ^b	Media ^c	s.e. ^b	Media
5 dpi	Control(1)	0,00045	0,00026	0,00198	0,00032	0,01709	0,00048	0,06250	0,00647	31,56
	Control (2)	0,00282	0,00047			0,02473	0,00178			
	FU25 (1)	0,00060	0,00029	0,00060	0,00029	0,05499	0,00364	0,05499	0,00364	91,65
	FU50 (1)	0,00050	0,00033	0,00050	0,00033	0,04993	0,00349	0,04993	0,00349	99,86
	FU100 (1)	0,00075	0,00033			0,00025	0,00013			
	FU100 (2)	0,00039	0,00028	0,00060	0,00018	0,00026	0,00015	0,00035	0,00010	0,58
	FU100 (3)	0,00070	0,00030			0,00049	0,00843			
10 dpi	Control(1)	0,00075	0,00037	0,00075	0,00037	0,00106	0,00058	0,00106	0,00058	1,41
	FU25 (1)	0,00021	0,00020	0,00021	0,00020	0,00065	0,00036	0,00065	0,00036	3,09
	FU50 (1)	0,00038	0,00025	0,00038	0,00025	0,01073	0,00116	0,01073	0,00116	28,2
	FU100 (1)	0,00000	0,00000			0,00120	0,00048			
	FU100 (2)	0,00060	0,00032	0,00020	0,00010	0,00148	0,00048	0,00147	0,00052	7,35
	FU100 (3)	0,00000	0,00000			0,00053	0,00032			
31 dpi	Control (1)	0,00050	0,00048	0,00050	0,00048	0,00050	0,00051	0,00050	0,00051	1
	FU100 (1)	0,00097	0,00037	0,00097	0,00037	0,00097	0,00038	0,00097	0,00038	1

^a Entre paréntesis se muestra el número correspondiente a la réplica biológica.^b Error estándar.^c Media calculada a partir de la d de las distintas réplicas para cada tratamiento y tiempo.^d Relación entre las d medias de MP y RdRp

Estos resultados ponen de manifiesto diferencias en la complejidad genética de las dos regiones de las cuasiespecies de TMV de muestras tratadas con 5-FU. Mientras que en la región RdRp se da una disminución moderada de la distancia genética, en la región MP tiende a aumentar a lo largo del tratamiento. Sin embargo, a tiempos largos sin FU la distancia genética no sólo se iguala en las dos regiones genómicas sino que en ambas la complejidad genética en las poblaciones que habían sido tratadas con 100 µg/ml de FU es el doble que en el control.

Como se ha indicado, el valor d de un espectro de mutantes representa una media de todas las distancias genéticas entre cualesquiera dos secuencias tomadas al azar, por lo que su coeficiente de variación (CV) da una idea del grado de dispersión de los valores que se promedian en el cálculo de la distancia genética. Si el CV es inferior a 1 estaría indicando que el espectro de mutantes analizado presenta poca heterogeneidad. Para comprobar si el 5-FU influye en el CV de las poblaciones virales, dividimos las muestras en dos grupos: muestras sin tratar y muestras tratadas (independientemente de la concentración de 5-FU), y comparamos el CV calculado para cada una de las regiones RdRp y MP. Tal como se muestra en la **Tabla 4** en las muestras control el CV de d en ambas regiones RdRp y MP fue cercano a 1 (0,963 y 1,330, respectivamente), lo que indica una amplia dispersión de datos de d (entre pares de secuencias) en torno a la media de la población, o lo que es lo mismo, que las secuencias que se incluyen dentro de la muestra difieren mucho de la media. En el caso de las muestras tratadas con 5-FU, en la región MP se observa una dispersión mayor de la distancia genética que en su control (1,904). Sin embargo, en el caso de la RdRp el CV descendió hasta 0,65 sugiriendo que en esta región en presencia de 5-FU se ha eliminado parte de la variabilidad de d que se observa en el control o en la región MP.

Tabla 4. Coeficiente de variación calculado para las regiones RdRp y MP de muestras control y muestras tratadas con 5-FU

CV	RdRp	MP
C-	0,963	1,330
5-FU	0,654	1,904

5. Heterogeneidad de las cuasiespecies víricas

Como medida de la heterogeneidad de las cuasiespecies víricas se calculó el índice de Shannon normalizado (S_n). Este índice da una medida de cuán diferente son los genomas en una distribución de mutantes teniendo en cuenta la frecuencia de cada secuencia en el espectro de mutantes. Los valores de este índice oscilan entre 0, si todos los genomas de la cuasiespecie son idénticos entre sí, y 1 en el caso de que todos fueran diferentes. Para su cálculo (indicado en el apartado 13.3. de Materiales y Métodos) se analizó toda la zona secuenciada en conjunto. Los resultados de la **Figura 5** indican que el S_n no mostró diferencias significativas (según ANOVA para un $p < 0,05$) entre tratamientos o tiempos. Sin embargo, sí mostró una alta variabilidad entre muestras, con valores entre 0,2 y 0,7, aunque la mayoría se situó en torno a valores cercanos a la media (0,5). Los valores máximos de entropía a 5 dpi los presentó una de las réplicas control, mientras que a 10 dpi fue una muestra de 100 $\mu\text{g/ml}$ de 5-FU.

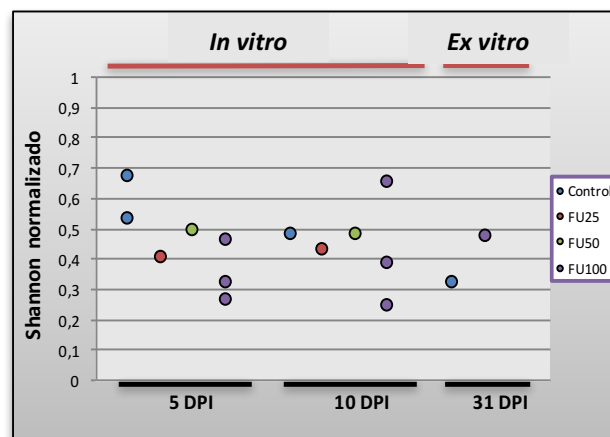


Figura 5. Índice de Shannon normalizado de los espectros de mutantes de TMV a 5, 10 y 31 dpi. El índice de Shannon se utilizó para medir la heterogeneidad del espectro de mutantes, empleando la fórmula $S_n = -[\sum_i (p_i \cdot \ln p_i)] / \ln N$ donde " p_i " es la frecuencia de cada secuencia en el conjunto de secuencias y " N " es el número total de secuencias analizadas. Los valores de S_n van de 0 (homogeneidad completa) a 1 (heterogeneidad completa).

6. Caracterización de las mutaciones en presencia de 5-FU

Para averiguar si el descenso de infectividad específica de las cuasiespecies tratadas con 100 $\mu\text{g/ml}$ de 5-FU a 10 dpi estaba relacionado con una actividad mutagénica del 5-FU, decidimos analizar en primer lugar el tipo de mutaciones halladas en las cuasiespecies

control y las tratadas con 100 µg/ml de 5-FU a 10 dpi. Los tipos y porcentajes respecto al total de mutaciones para cada una de las regiones analizadas RdRp y MP, así como para la zona secuenciada en conjunto (total) se muestran en la **Tabla 4**.

Tabla 4. Tipos de mutaciones encontradas en los espectros de mutantes de TMV a 10 dpi.

Tipos de mutaciones	CONTROL ^a			FU100 ^a		
	RdRp	MP	TOTAL	RdRp	MP	TOTAL
Transiciones	%	%	%	%	%	%
A→G	0,0	0,0	0,0	33,3	15,0	17,4
G→A	0,0	40,0	22,2	0,0	20,0	17,4
C→U	25,0	20,0	22,2	0,0	10,0	8,7
U→C	50,0	20,0	33,3	66,7	30,0	34,8
Total	75,0	80,0	77,8	100,0	75,0	78,3
Transversiones						
A→C	0,0	0,0	0,0	0,0	0,0	0,0
C→A	0,0	0,0	0,0	0,0	5,0	4,3
A→T	0,0	0,0	0,0	0,0	0,0	0,0
T→A	0,0	0,0	0,0	0,0	0,0	0,0
C→G	0,0	0,0	0,0	0,0	0,0	0,0
G→C	0,0	0,0	0,0	0,0	5,0	4,3
G→T	0,0	20,0	11,1	0,0	0,0	0,0
T→G	25,0	0,0	11,1	0,0	5,0	4,3
Total	25,0	20,0	22,2	0,0	15,0	13,0
Sustituciones	100	100	100	100	90,0	91,3
Inserciones	0,0	0,0	0,0	0,0	5,0	4,3
Deleciones	0,0	0,0	0,0	0,0	5,0	4,3
Total	100	100	100	100	100	100
No sinónimas						
Cambio de sentido	50	60	55,55	66,6	45	47,8
Sin sentido	0	0	0	0	0	0
Cambio marco de lectura	0	0	0	0	10	8,6
Sinónimas	50	40	44,4	33,3	45	43,4
Total	100	100	100	100	100	100

^aMutaciones de los espectros de mutantes de las réplicas 1 y 2 control (88961 nt secuenciados) y réplicas 1, 2 y 3 de 100 µg/ml de 5-FU (82881 nt secuenciados) descritas en la Tabla A1 del Anexo

A 10 dpi se identificaron un total de 32 mutaciones, entre muestras control y tratadas con 5-FU, de las cuales más del 75% fueron transiciones. Tan sólo se encontró una deleción así como una inserción en la región MP de las muestras tratadas con 5-FU. En la región RdRp del control se encontraron transiciones C→U (25%) y U→C (50%) y

transversiones T→G (25%) mientras que en las muestras tratadas con 5-FU sólo aparecieron las transiciones U→C (66,7%) y A→G (33,3%), que son los dos tipos de transiciones más comúnmente asociadas al efecto del FU. En la región MP en las muestras tratadas con 5-FU se detectó un incremento de las transiciones A→G y U→C, y también aparecieron transversiones C→A, G→C y T→G aunque en bajo porcentaje. Considerando la región secuenciada en conjunto en las muestras tratadas con el análogo a 10 dpi hubo un aumento de transiciones A→G, pero no de U→C, con respecto al control. El aumento de transiciones debidas a la incorporación del 5-FU en el RNA vírico no se registró a 5 dpi (ver **Tabla A2** en Anexo).

A continuación realizamos un análisis de la relación entre las transiciones asociadas a 5-FU y las no asociadas al análogo para las cuasiespecies control y las tratadas con 25, 50 y 100 µg/ml de 5-FU a 10 dpi mediante la fórmula:

$$[(A \rightarrow G) + (U \rightarrow C)] / [(G \rightarrow A) + (C \rightarrow U)]$$

Los resultados mostraron un incremento de las transiciones A → G y U → C en las muestras de TMV de plantas tratadas con 5-FU, siendo este ratio 1,5 veces mayor en las muestras de 25 µg/ml (ratio=1) con respecto al control (ratio=0,667) y de 2,24 veces superior (ratio=1,5) en el caso de las muestras tratadas con 50 y 100 µg/ml con respecto a las muestras sin tratar. Este incremento de las mutaciones asociadas al análogo de base relaciona directamente la actividad mutagénica del 5-FU con la concentración que se utilizó para cada tratamiento.

Las mutaciones no sinónimas fueron en general más abundantes que las sinónimas, especialmente en la región RdRp de la muestra tratada con 100 µg/ml de 5-FU, en las que sólo un 33,3% de las mutaciones encontradas eran sinónimas (**Tabla 4**).

Los valores de aceptabilidad de los cambios de aminoácido empleando la matriz de estructura genética propuesta por Feng y colaboradores (Feng et al., 1984) para las mutaciones halladas en las cuasiespecies control y las tratadas con 25, 50 y 100 µg/ml de 5-FU, a 5 y 10 dpi, se muestran en la **Figura 6**. A 10 dpi el número de mutaciones no sinónimas (con valores de aceptabilidad inferiores a 6) disminuye en todas las muestras, control o tratadas con 5-FU. Estos resultados sugieren una mayor exploración del espacio de secuencia de todas las cuasiespecies de TMV en los días iniciales de infección, favoreciéndose tanto cambios de aminoácido más drásticos como sinónimos. Sin

embargo, a 10 dpi en el control se permitirían menos cambios drásticos que en cualquiera de los tratamientos con 5-FU.

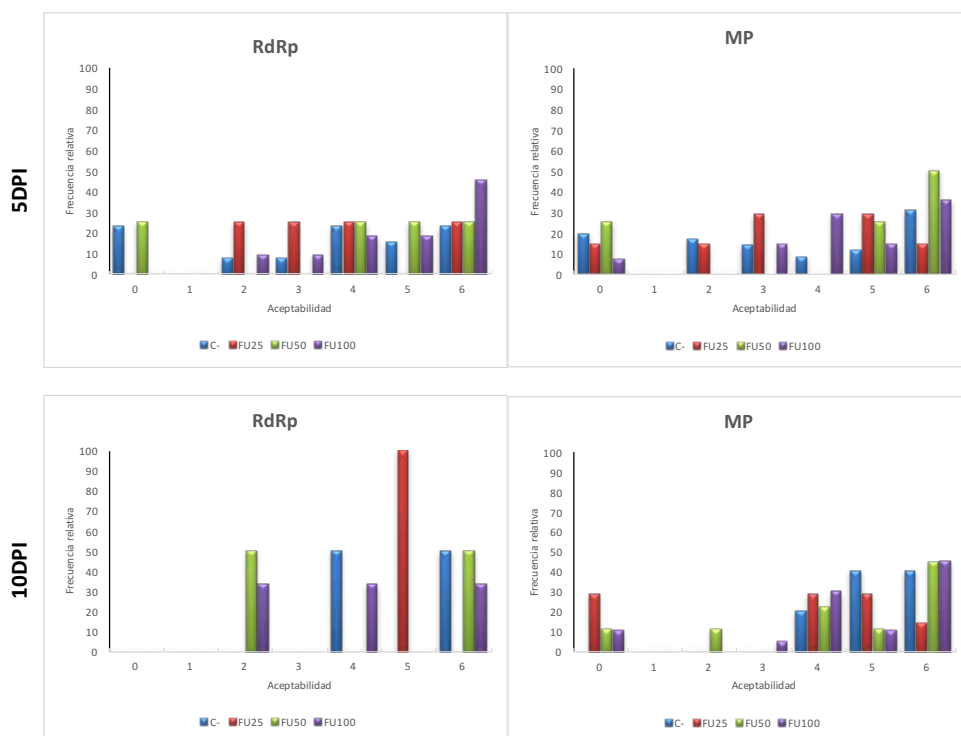


Figura 6. Aceptabilidad de los cambios de aminoácido de las cuasiespecies de TMV. En ordenadas se representa la frecuencia relativa respecto al número total de cambios de aminoácido de las cuasiespecies de TMV control (Control) y las tratadas con 25, 50 y 100 µg/ml de 5-FU a 5 y 10 dpi. En abscisas se representan los valores obtenidos para cada cambio de aminoácido según la matriz SG (Feng et al., 1983) en una escala de 1 a 6, siendo 1 el cambio más drástico y 6 el menos drástico (mutación sinónima). El valor 0 corresponde a cambios hacia codones de terminación debidos a mutaciones sin sentido, no contemplados en la matriz SG

7. Análisis del espacio de secuencia

Como se indicó anteriormente, se ha estudiado el espectro de mutantes de 15 cuasiespecies de TMV tratadas a diferentes concentraciones de 5-FU a diferentes tiempos. Para determinar si el 5-FU influye en el espacio de secuencia que exploran las cuasiespecies y si hay alguna relación con la concentración de 5-FU y/o la duración del tratamiento, se estudió la distribución de las mutaciones encontradas a 5 y 10 dpi en la zona secuenciada. La **Figura 7** muestra la distribución de las mutaciones detectadas en la zona secuenciada del genoma de TMV a 5 (panel A) y 10 dpi (panel B). Para evaluar si la distribución de estas mutaciones era aleatoria realizamos un Test no paramétrico de

rachas (Wald and Wolfowitz, 1943) de las posiciones mutadas diferenciando entre ORFs y la presencia o ausencia del análogo de base. Dicho test determina si las rachas o sucesiones de mutaciones–no mutaciones siguen un orden de aparición aleatorio o no en la secuencia estudiada. El Test de rachas también se empleó para averiguar si las frecuencias de mutación encontradas en las dos regiones diferían entre sí a ambos tiempos de tratamiento.

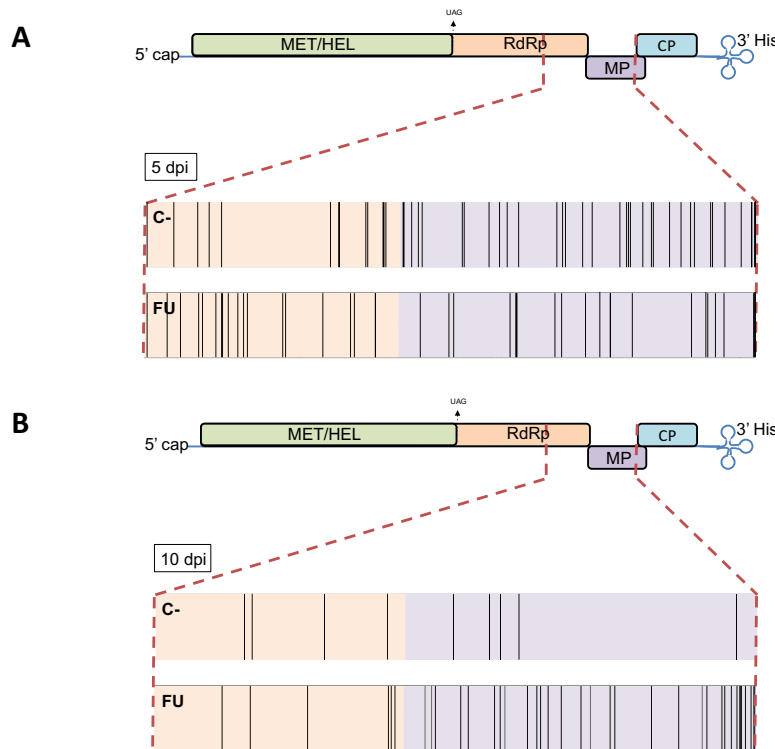


Figura 7. Localización de las mutaciones puntuales presentes en las regiones RdRp y MP secuenciadas. Esquema mostrando la ubicación de las mutaciones puntuales observadas en muestras sin tratar (Control) y muestras tratadas con 100 µg/mL de 5-FU (FU100) a 5 dpi(A) y 10 dpi (B). En la parte superior de cada panel se representa el esquema del RNA de TMV completo, indicando el cap en el extremo 5', sus cuatro ORF, metiltransferasa/helicasa (MET/HEL), RNA polimerasa dependiente de RNA (RdRp), proteína del movimiento (MP) y proteína de la cápside (CP) y la estructura RNAt_{his} en el extremo 3'. La región secuenciada se señala con líneas discontinuas y se muestra ampliada en la parte inferior. Las mutaciones están representadas como líneas negras sobre fondo rosa en la región RdRp o fondo morado en la MP.

Tal como se muestra en la **Tabla 5** en primer lugar se analizó la distribución de adeninas y uracilos en las regiones de RdRp y MP y se comprobó que era aleatoria en ambas regiones. Tas el análisis de las mutaciones, la región RdRp presentó una distribución no aleatoria significativa ($p < 0,01$) en el control tanto a 5 como a 10 dpi, mientras que en presencia de 5-FU dicha distribución sí fue aleatoria ($p > 0,05$). Sin embargo en la región

MP en las muestras control se observó una distribución de mutaciones aleatoria tanto a 5 como a 10 dpi, mientras que en presencia de 100 µg/ml 5-FU la distribución de las mutaciones no fue aleatoria a 5 pero sí a 10 dpi según el Test de rachas realizado ($p < 0,01$).

Tabla 5. Test de rachas para determinar la distribución o no aleatoria de las mutaciones en RdRp y MP.

	5 dpi		10 dpi	
	RdRp	MP	RdRp	MP
A/U^a	0,969 ^b	0,980	0,969	0,980
Control^c	0,014	0,690	0,002	1,000
FU100^c	0,535	0,001	0,557	0,484

^aDistribución de adeninas y uracilos en las regiones de RdRp y MP para comprobar la distribución aleatoria de éstas en ambas regiones.

^bEn el Test de rachas los valores inferiores a 0,05 indican una distribución no aleatoria de las mutaciones y los valores superiores a 0,05 indican distribución aleatoria de mutaciones.

^cMutaciones de los espectros de mutantes Control (réplica 1) y 100 µg/ml de 5-FU (réplicas 1, 2 y 3), todas a 10 dpi, descritas en la Tabla A1 (ver Anexo).

Estos resultados ponen de manifiesto diferencias entre regiones con respecto a las posiciones nucleotídicas que exploran las cuasiespecies control así como el efecto que tiene el análogo sobre dicha exploración. Así, en la región RdRp existe restricción al cambio en ciertas posiciones nucleotídicas que se pierde en presencia de 5-FU. Por el contrario, en la región MP se permite una mayor exploración del espacio de secuencia que no se ve alterada por la presencia del análogo de base a 10 dpi.

8. Dirección de las fuerzas selectivas

Dado que las cuasiespecies *in vivo* son conjuntos de secuencias genómicas muy relacionadas entre sí generadas por un proceso de mutación-selección (Domingo et al., 2012) nos propusimos averiguar qué tipo de selección está actuando en las diferentes poblaciones de TMV durante la mutagénesis con 5-FU. Para ello calculamos la diversidad de sustituciones sinónimas por sitio sinónimo (d_s) y la diversidad de sustituciones no sinónimas por sitio no sinónimo (d_{ns}) así como el ratio d_{ns}/d_s en las muestras control y tratadas con 100 µg/ml de 5-FU tanto a 5 como a 10 dpi. Para ello se empleó el método Pamilo-Bianchi-Li (Li, 1993; Pamilo and Bianchi, 1993) mediante el software MEGA6 con un *bootstrap* de 1000. La razón d_{ns}/d_s proporciona una estimación del grado y la dirección de la selección natural que actúa en las regiones codificantes del

genoma viral. En la **Tabla 6** se muestran los valores d_{ns}/d_s para cada cuasiespecie control y tratada con 5-FU a 5 y 10 dpi para las regiones RdRp y MP. Los resultados mostraron nuevamente diferencias entre las dos regiones analizadas. En la región de la RdRp a los 5 días en el control el valor $d_{ns}/d_s = 1$ (1,05), lo que indicaría que selección es neutra, para pasar a los 10 dpi a una selección negativa o purificadora ($d_{ns}/d_s < 1$; 0,44). Sin embargo, en presencia de 100 µg/ml de 5-FU la selección comenzaría siendo negativa (0,25) para pasar a ser neutra (1,10) a 10 dpi. En el caso de la región MP al comienzo de la infección en ambos casos la selección sería negativa pero a 10 dpi en ausencia del análogo de base existiría una selección positiva, dado que el ratio $d_{ns}/d_s > 1$ (1,53), mientras que en presencia de 5-FU la selección sería neutra (0,82).

Tabla 6. Análisis de la diversidad en posiciones no sinónimas (d_{ns}) y sinónimas (d_s) en las regiones RdRp y MP en muestras control y tratadas con 100 µg/ml de 5-FU a 5 y 10 dpi

		RdRp				
		d_{ns}	s.e. ^a	d_s	s.e.	d_{ns}/d_s
5dpi	Control	0,00051	0,00020	0,00047	0,00026	1,08511
	FU100	0,00033	0,00015	0,00129	0,00058	0,25581
10 dpi	Control	0,00049	0,00034	0,00109	0,00070	0,44954
	FU100	0,00021	0,00014	0,00019	0,00019	1,10526

		MP				
		d_{ns}	s.e.	d_s	s.e.	d_{ns}/d_s
5dpi	Control	0,00108	0,00024	0,00139	0,00046	0,77698
	FU100	0,00027	0,00012	0,00047	0,00024	0,57447
10 dpi	Control	0,00117	0,00054	0,00076	0,00054	1,53947
	FU100	0,00144	0,00094	0,00173	0,00094	0,83237

^a Error estándar

Estos resultados muestran que la selección natural operaría en direcciones opuestas en las dos regiones en ausencia de análogo. En presencia de 100 µg/ml de 5-FU en ambas regiones la selección actúa de la misma manera, es decir, a tiempos tempranos en ambas domina la selección negativa y a 10 dpi se hace neutra en las dos. Esto sugiere que el 5-FU *in vivo* altera el equilibrio mutación–selección de ambas regiones anulando la dirección de la selección.

9. Influencia del FU en la distribución de la variabilidad de TMV

Para determinar de forma global si el FU ejerce algún efecto sobre la distribución de la variabilidad genética en cada una de las dos regiones genómicas se utilizó el test de análisis de la varianza molecular (*Analysis of MOlecular VAriance*, AMOVA) (Excoffier et al., 1992) mediante el programa Arlequin 3.1. Este test se utiliza para evaluar la distribución de la variabilidad genética y se basa en el tipo y frecuencia de todas las secuencias de todas las muestras y en la posibilidad de que dicha diversidad sea aleatoria. Tal como se muestra en la **Tabla 7**, para cada región genómica analizada (RdRp y MP) se estudió el porcentaje de variación a tres niveles diferentes, variación entre tratamientos, entre muestras y entre haplotipos, entendiendo haplotipo como cada una de las secuencias diferentes existentes dentro de una cuasiespecie, independientemente de su frecuencia dentro de esta. En la región RdRp los resultados muestran claramente que la mayor variación se encontró al comparar el grupo control con las muestras tratadas con 5-FU, alcanzando un porcentaje de variación del 52,85%. La variación encontrada entre haplotipos fue del 35,32% y entre muestras biológicas como réplicas independientes, tan sólo del 11,82%. Estos resultados indican que la distribución de la variabilidad genética de las cuasiespecies en esta región está muy influida por los tratamientos con 5-FU. En el caso de la región MP **Tabla 8**, al contrario que lo que ocurrió en la RdRp sólo se observó variación al comparar haplotipos lo que indica una distribución aleatoria de la variabilidad genética de las muestras tanto en presencia como en ausencia de tratamiento con FU.

Tabla 7. Análisis molecular de la varianza (AMOVA) para el estudio de la variabilidad de haplotipos en la región RdRp de TMV

Fuente de variación	g.l.	Suma de cuadrados	Varianza muestral	Porcentaje de variación	Probabilidad
Entre tratamientos ^a	1	25,783	0,19905	52,85 %	P<0,05
Entre muestras ^b	10	11,311	0,04453	11,82 %	P<0,05
Entre haplotipos ^c	268	35,652	0,13303	35,32 %	P<0,05
Total	279	72,746	0,37660		

^a Muestras Control frente a muestras tratadas (se comparan 2 valores)

^b Control, 25, 50 y 100 µg/ml de 5-FU a 5 y 10 dpi (se comparan 11 valores entre sí)

^c variabilidad entre secuencias (se comparan 269 valores entre sí)

Tabla 10. Análisis molecular de la varianza (AMOVA) para el estudio de la variabilidad de haplotipos en la región MP de TMV

Fuente de variación	g.l.	Suma de cuadrados	Varianza muestral	Porcentaje de variación	Probabilidad
Entre tratamientos	1	0,224	-0,00377	<1%	P>0,05
Entre muestras	11	6,708	0,01190	3,42 %	P<0,05
Entre haplotipos	292	99,383	0,34035	97,67 %	P<0,05
Total	304	106,315	0,34849		

^a Muestras Control frente a muestras tratadas (se comparan 2 valores)

^b Control, 25, 50 y 100 µg/ml de 5-FU a 5 y 10 dpi (se comparan 11 valores entre sí)

^c variabilidad entre secuencias (se comparan 269 valores entre sí)

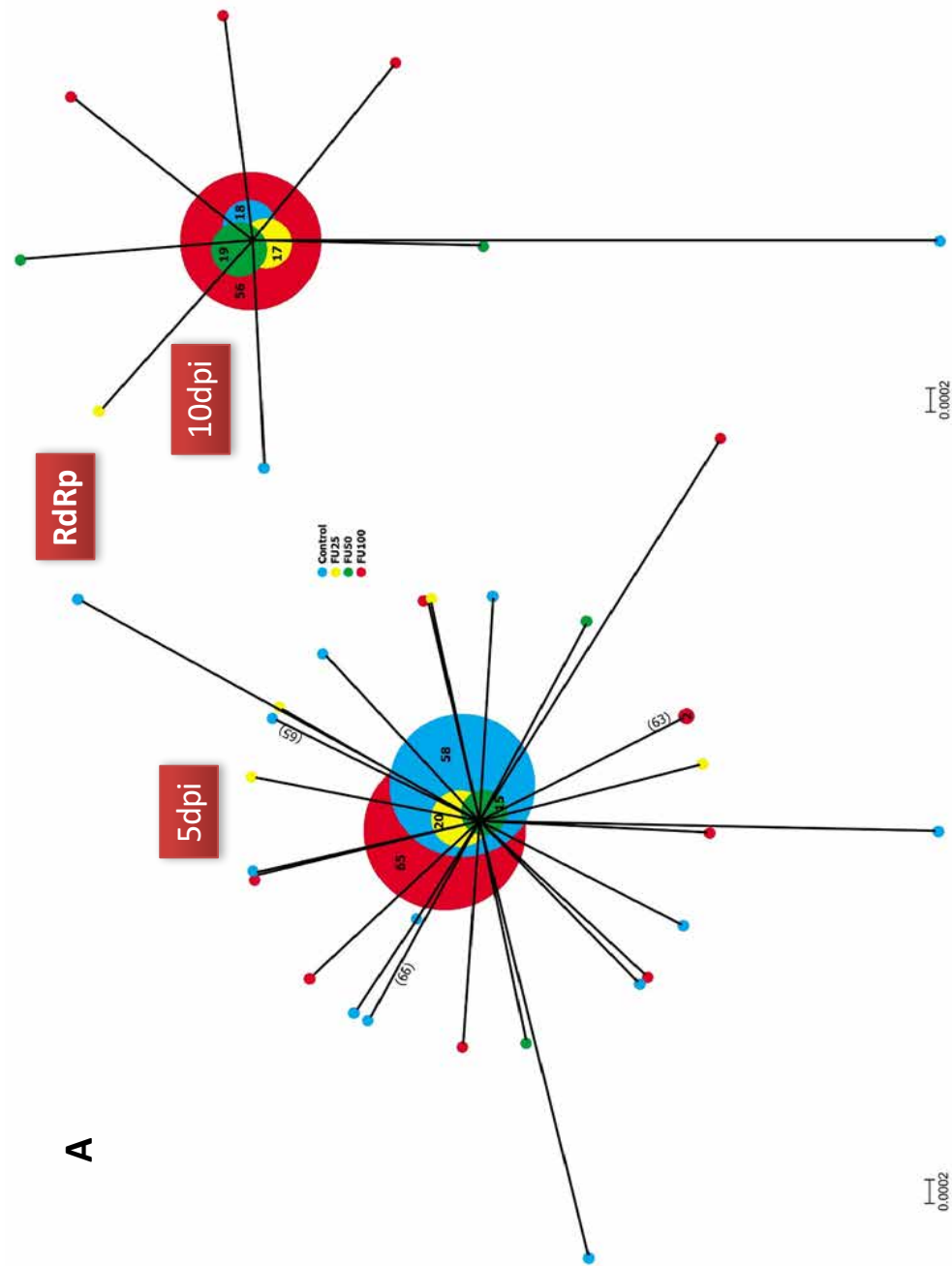
Estos resultados refuerzan la idea de que existen claras diferencias en el efecto mutagénico del 5-FU *in vivo* según la región analizada, lo que puede estar relacionado con la función de las proteínas que codifican dichas regiones.

10. Análisis filogenético del espectro de mutantes

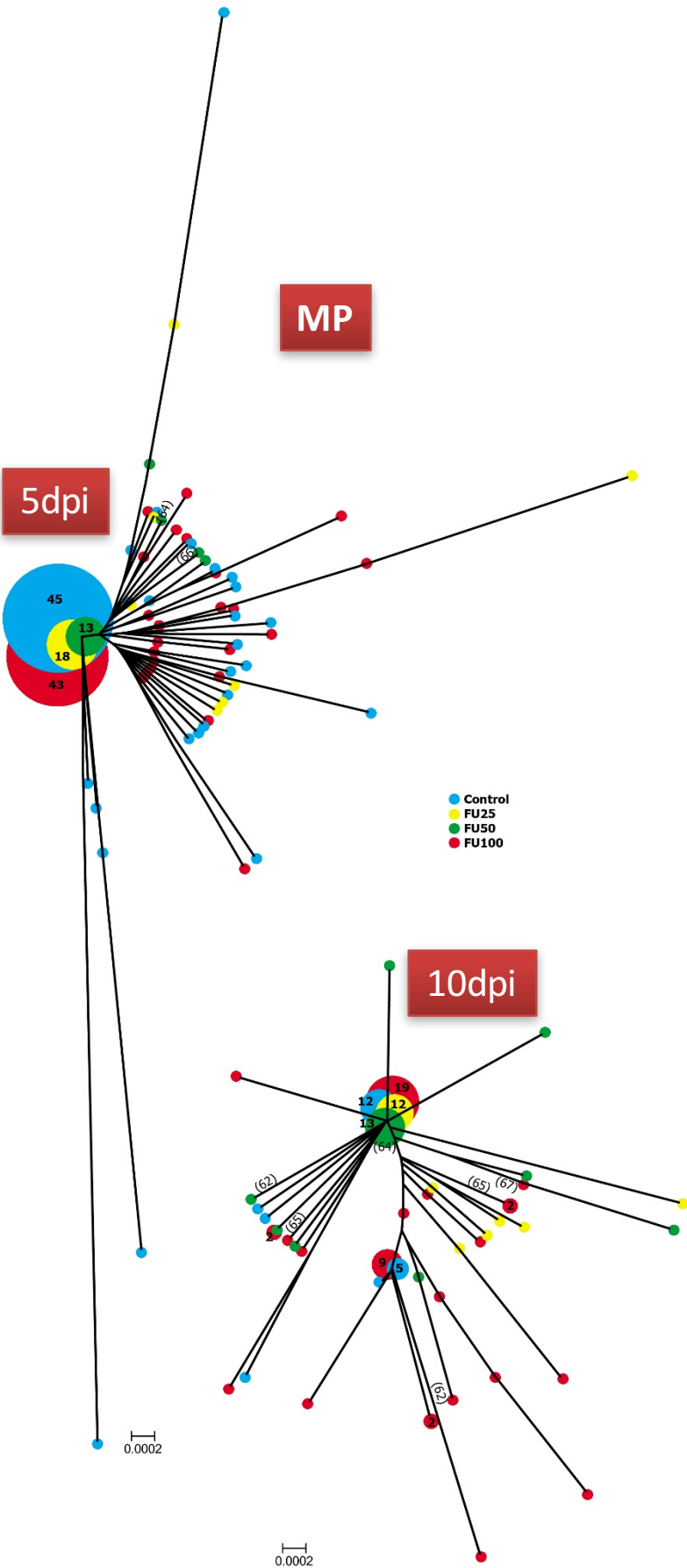
Las relaciones filogenéticas de las secuencias de las regiones genómicas RdRp y MP se obtuvieron a partir de alineamientos múltiples de secuencias mediante el método de *neighbor-joining* con un *bootstrap* de 1000. La **Figura 8** muestra un árbol sin raíz que

representa la filogenia para cada una de las dos regiones genómicas a 5 y 10 dpi. La composición, frecuencia y relación filogenética de las secuencias dentro de cada una de las cuasiespecies analizadas determinó las diferentes formas de los árboles. Los árboles con mayor dispersión, tanto a 5 como a 10 dpi, corresponden a la región MP (Panel B) mientras que la menor ramificación se da en la región RdRp a 10 dpi (Panel A). En esta región sólo se observó una secuencia maestra en cada una de las cuasiespecies analizadas tanto control como con 5-FU. Sin embargo, en la región MP a 10 dpi pudo observarse la existencia de una segunda secuencia maestra compartida por muestras control y de 100 µg/ml de 5-FU. El análisis filogenético a 5 dpi en esta región mostró una dispersión homogénea de las secuencias con independencia del tratamiento, mientras que a 10 dpi se observó una mayor diversificación de las secuencias pertenecientes a 100 µg/ml de 5-FU. Estos resultados ilustran las diferencias en los espectros de mutantes entre regiones genómicas observadas anteriormente en este capítulo.

Figura 8. Relación filogenética entre secuencias de los espectros de mutantes de TMV para las regiones RdRp (A) y MP (B). Para la reconstrucción filogenética se empleó el método *neighbor-joining* con un *bootstrap* de 1000 réplicas bajo el modelo de sustitución de máxima verosimilitud implementado en el programa MEGA6 (Tamura et al 2013). Los valores de *bootstrap* superiores a 60 están indicados entre paréntesis. Están representadas todas las secuencias de los espectros de mutantes de TMV control y tratadas con 25, 50 y 100 µg/ml de 5-FU a 5 y 10 dpi incluidas en la Tabla A1 (Anexo). Cada secuencia individual está representada por un círculo con diferente color según el tratamiento. El tamaño de los círculos es proporcional al número de secuencias idénticas y su número está indicado con un valor dentro del mismo.



B



11. Determinación de ribonucleótidos trifosfato (NTPs) y FUTP intracelulares

El desbalance en los niveles de nucleótidos intracelulares es mutagénico. Para conocer si el tratamiento con 5-FU *in vivo* alteró los niveles de NTPs intracelulares determinamos su concentración en las plantas infectadas mediante HPLC. Para ello, llevamos a cabo extracciones de nucleótidos a partir de plantas *N. tabacum* cv Samsun nn crecidas *in vitro* en ausencia de FU y en presencia de 100 µg/ml de 5-FU durante 5 y 10 días o tras 10 días con 5-FU *in vitro* seguido de 21 días sin análogo en maceta (31 dpi) tal como se muestra en el apartado 6.1. de Materiales y Métodos. Las extracciones se analizaron en columna de intercambio aniónico y los cuatro NTPs celulares se determinaron por comparación con sus respectivos estándares. Los resultados mostrados en la **Figura 9** indican que no existieron desbalances en los niveles de NTPs a ningún tiempo ni en presencia ni en ausencia de 100 µg/ml de 5-FU para un $p < 0,05$.

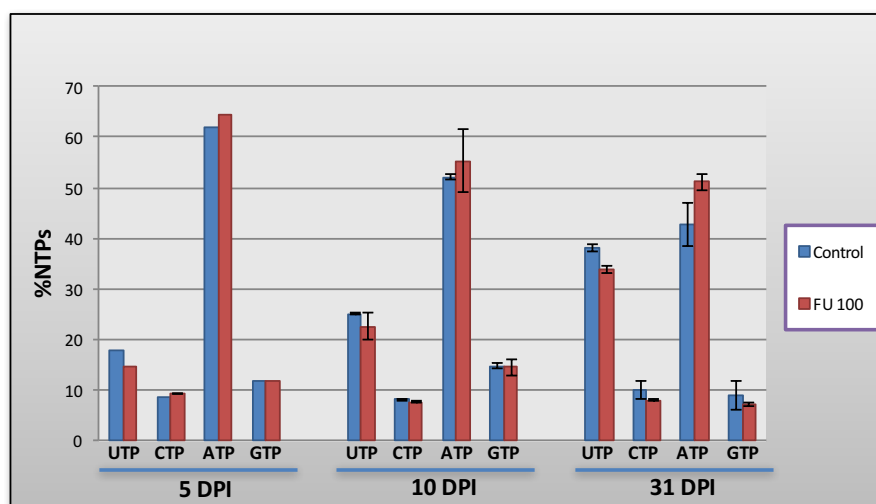


Figura 9. Determinación de ribonucleótidos trifosfato en plantas de *N. tabacum* cv Samsun nn infectadas con TMV. Se representa el porcentaje de CTP, UTP, ATP y GTP con respecto al total de los cuatro nucleótidos presentes en extractos de plantas infectadas con TMV en ausencia o presencia de 100 µg/ml de 5-FU tras 5 y 10 días de tratamiento *in vitro*, o tras 10 días de dicho tratamiento seguido de 21 días en ausencia de 5-FU *ex vitro*. Los resultados corresponden a tres réplicas técnicas por cada muestra biológica en cada tratamiento a 5 dpi, y a tres muestras biológicas con tres réplicas técnicas a 10 y 31 dpi. Las barras de error representan la desviación estándar de las réplicas técnicas.

Por otro lado no fue posible detectar las formas metabolizadas del 5-FU, ni el FUDP ni el FUTP, ni siquiera aumentando la cantidad de muestra inicial cuatro veces y obteniendo concentraciones de NTPs del orden de nmoles. A raíz de estos resultados, decidimos no

analizar por HPLC las muestras procedentes de plantas tratadas con concentraciones menores de 5-FU. No obstante, el análisis por HPLC de la concentración de 5-FU en el medio de cultivo durante los 10 días de tratamiento indicó que esta se mantuvo para cada tratamiento (resultados no mostrados).

Dado que no fuimos capaces de detectar el FUTP en las muestras tratadas con el análogo, se comprobó si las células de *N. tabacum* incorporaban el 5-FU al metabolismo de las pirimidinas para sintetizar FUTP. Para ello, realizamos un estudio de la cinética de metabolización del 5-FU empleando cultivos de células BY-2 de *N. tabacum*. En este experimento se crecieron en cultivo líquido células indiferenciadas BY-2 en ausencia o presencia de 25, 50 y 100 µg/ml de 5-FU. Se tomaron muestras de cada cultivo al inicio y tras 24, 48, 72 y 96 horas postratamiento y se determinaron los niveles de 5-FUTP y de los cuatro NTPs por HPLC. Como se observa en la **Figura 10**, en las muestras de los cultivos de células BY-2 tratadas con 5-FU se detectó el FUTP desde las 24 horas de tratamiento, alcanzando niveles de 21, 61 y 460 pmoles en presencia de 25, 50 y 100 µg/ml de 5-FU respectivamente, desapareciendo a las 72 h. Sin embargo, en las células tratadas con 100 µg/ml 5-FU el FUTP intracelular se mantuvo en concentraciones cercanas a 600 pmoles desde las 24 a las 48 horas postratamiento y dejó de detectarse a las 96 horas. Los resultados demuestran que en cultivos celulares de *N. tabacum*, el análogo puede incorporarse al metabolismo de las pirimidinas. Adicionalmente, estos resultados también mostraron que en cultivos de células BY-2, el tratamiento con dosis de 100 µg/ml de 5-FU disminuyó la cantidad de CTP intracelular, que recuperó valores del control cuando el FUTP desaparece (**Figura 10**). La concentración de UTP, ATP y GTP permanece sin grandes variaciones a lo largo de todo el estudio independientemente de la concentración de 5-FU que se haya aplicado al cultivo de BY-2.

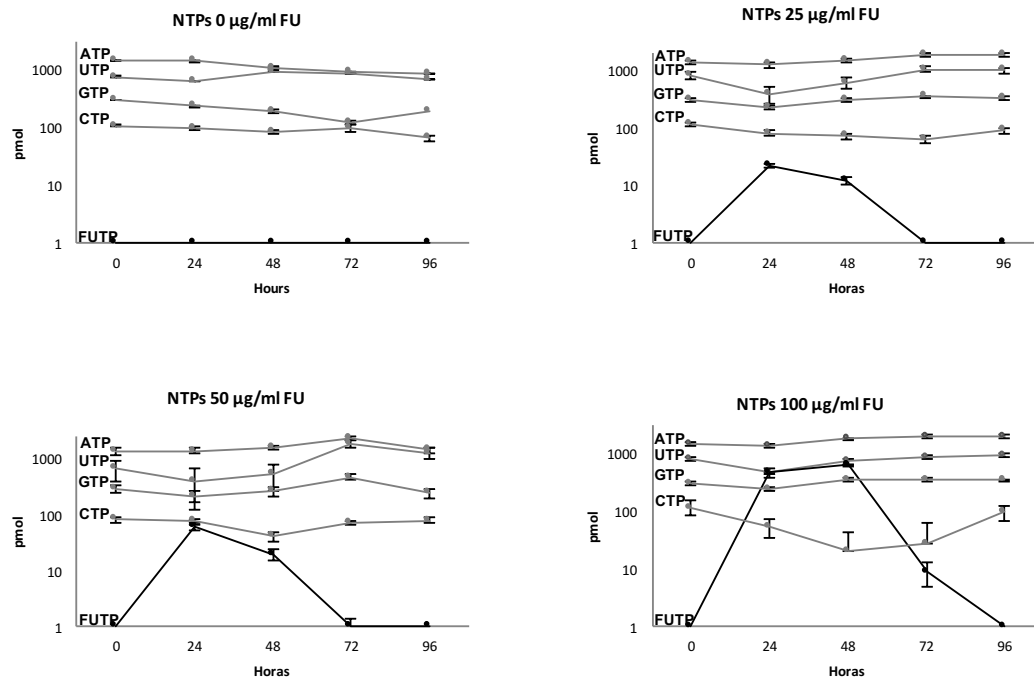


Figura 10. Determinación de la concentración de 5-FU trifosfato (FUTP) y de los ribonucleótidos trifosfato intracelulares en células BY-2. Cultivos de células BY-2 a una densidad inicial de 10^6 céls/ml se incubaron en condiciones de oscuridad a 27°C en agitación en ausencia (Control) o en presencia de 25, 50 y 100 µg/mL de 5-FU. A los tiempos indicados se tomaron muestras y se analizó el contenido intracelular de ribonucleótidos trifosfato ATP, UTP, GTP y CTP y FUTP mediante HPLC en columna de intercambio aniónico Partisil SAX como se indica en Materiales y métodos. Las barras de error representan la desviación estándar de 4 réplicas biológicas para cada tratamiento.

Discusión

La mutagénesis letal es una nueva terapia antiviral cuyo fundamento es el aumento de la tasa de error de los virus RNA mediante el uso de agentes mutagénicos. Análogos de base como el 5-FU o de nucleósido como la 5-hidroxi-desoxicitidina, la ribavirina y la 5-azacitidina, entre otros, presentan actividad antiviral asociada a mutagénesis incrementada en cultivos celulares frente a una amplia gama de virus RNA. El aumento de las frecuencias de mutación van acompañada de una disminución la infectividad vírica lo que conduce en algunos casos a la extinción completa de la población del virus (Domingo et al., 2012). En la actualidad, los mecanismos moleculares que conducen a la extinción virus mediante mutagénesis incrementada no están totalmente dilucidados. Existe evidencia experimental apoyando el modelo de defección letal de los virus de ARN, en el que debido a la mutagénesis se generan virus mutantes con un bajo grado de mutación, denominados defectores, capaces de replicarse y que conducen al virus infeccioso hacia la extinción.

Mientras que la mayoría de los estudios sobre mutagénesis letal se han centrado en cultivos de células de virus de ARN de animales pocos estudios lo han abordado *in vivo*. En este capítulo de tesis hemos estudiado la mutagénesis letal de un virus RNA de plantas para comprender las bases moleculares de la pérdida de infectividad *in vivo*. Para ello hemos empleado el análogo de base 5-FU en plantas de *N. tabacum* ya que estudios previos documentaban disminución de la infectividad de TMV en este hospedador (Davern and Bonner, 1958; Gleason and Fraenkel-Conrat, 1976). El tratamiento se realizó un día antes de la inoculación del virus para que el metabolito principal derivado del 5-FU, el FUTP, necesario para que el 5-FU se pueda incorporar al RNA vírico, estuviera presente en la fase activa de replicación del virus. Al analizar el contenido de ribonucleósidos trifosfato intracelular de las plantas tratadas con el análogo durante 10 días no detectamos desbalances de los mismos, lo que descarta una actividad mutagénica derivada de la no disponibilidad de alguno de los ribonucleótidos. Pero tampoco fuimos capaces de detectar el FUTP, ni siquiera cuando aumentamos la cantidad de tejido de planta hasta obtener cantidades del orden de nmoles para los NTPs. La falta de detección el FUTP podía deberse bien a que las células *N. tabacum* no metabolizasen el 5-FU o bien a una baja disponibilidad de FUTP en las células, por debajo

del nivel de detección de nuestro método. Lo primero quedó descartado con el ensayo en cultivos celulares BY-2 de *N. tabacum* ya que, usando el mismo protocolo de extracción, conseguimos detectar el FUTP en la fracción intracelular a las 24 h de la adición de 5-FU. Lo segundo es complicado de determinar dado que desconocemos qué porcentaje de 5-FU puede llegar a las hojas y cuál es la tasa de síntesis y de degradación de FUTP en ellas. Špak y colaboradores (Špak et al., 2011) en un estudio acerca de la actividad antiviral del fosfonato nucleósido acíclico tenofovir sobre el pararretrovirus *Cauliflower mosaic virus* (CaMV) en plantas de *Brassica pekinensis* crecidas *in vitro*, analizaron la cantidad de las formas fosforiladas del análogo en las hojas. A las 96 h de añadir el análogo tritiado detectaron, mediante un protocolo similar al nuestro por HPLC, la máxima concentración de los derivados equivalentes a los NDPs y NTPs. A pesar de que las formas fosforiladas sólo representaron el 4,5 % del nucleósido añadido a los cultivos, el tenofovir a 50 µg/ml redujo significativamente el título de CaMV en 3-9 semanas. Este estudio indicó que el virus es sensible a cantidades muy bajas de las formas metabolizadas de análogo, aunque se desconoce si el modo de acción del mismo puede incluir la mutagénesis.

En nuestro experimento a los 5 días de tratamiento no se detectó actividad antiviral a ninguna de las dosis empleadas, sin embargo, a los 10 días con la dosis de 100 µg/ml de 5-FU la infectividad de TMV disminuyó un 72%. Aunque el descenso fue moderado y significativo sólo para la dosis más alta de 5-FU, observamos una tendencia a una reducción de la infectividad de TMV al aumentar la dosis de FU. Es probable que un tratamiento más prolongado, como en el caso del estudio con tenofovir comentado, o con una cantidad de mutágeno más elevada la reducción de la infectividad hubiera sido más notable o incluso llegado a eliminar el virus.

La carga viral, medida como moléculas genómicas de TMV, no se vio afectada con el tratamiento de 5-FU a ninguna dosis. Sin embargo la infectividad específica del virus disminuyó en las muestras tratadas con 100 µg/ml de 5-FU. Estos resultados concuerdan con lo encontrados por Barta y colaboradores (Barta et al., 1981) para otro análogo de uracilo, el tiouracilo (TU) que, en protoplastos de tabaco causó una reducción de un 10-20% de la infectividad específica de TMV. Sin embargo, el TU no inhibió la incorporación de ³²P en el RNA de TMV ni tampoco la multiplicación viral tanto en infecciones sincronizadas de protoplastos de tabaco como en protoplastos extraídos de hojas

infectadas con TMV. Dado que el número de células infectadas en hojas tratadas con TU disminuía en un 60 a un 70% con respecto a las hojas no tratadas, estos autores sugirieron que en las hojas el TU podía expresar su efecto inhibitorio sin impedir la inhibición directa de la multiplicación viral, generando partículas defectivas de TMV incapaces de infectar célula a célula. La disminución de la infectividad específica es una de las características distintivas del modelo de defeción letal que explica los acontecimientos moleculares que subyacen a la mutagénesis letal.

Otra de las características distintivas de la mutagénesis letal mediada por defectores es la invarianza de la secuencia consenso. En nuestro caso se ha cumplido en todas las muestras excepto en una de las cuasiespecies de 100 µg/ml de FU a los 10 dpi (réplica 1). En ella la mutación G5626A se halló en 16 de los 20 clones moleculares analizados, en alguno de ellos asociado a otras mutaciones, lo que indica una alta heterogeneidad en esa posición en los espectros de mutantes. La misma mutación apareció en la muestra control a 10 dpi en 5 de 20 clones, en este caso como única mutación. El cambio de aminoácido que causa, Asp242Asn, tiene un valor de aceptabilidad de 5 (poco drástico) en la matriz SG de Feng y colaboradores (Feng et al., 1984) y está localizada en la MP. En esa región el test de rachas encontró que a los 10 dpi las mutaciones se distribuían aleatoriamente, independientemente del tratamiento con 5-FU. En ausencia de análogo encontramos que la selección era positiva mientras con 100 µg/ml de 5-FU la selección pasaba a ser neutra. Podemos especular que la mutación G5626A sea selectivamente neutra y su imposición, aunque no total en la secuencia consenso, sea debida a la deriva genética debido a los frecuentes cuellos de botella poblacionales a los que se enfrenta el virus y que han sido descritos para TMV en este hospedador y que pueden disminuir su censo a unidades de viriones (Sacristán et al., 2003). En apoyo a esto está el hecho de que dicha mutación esté también presente en una muestra control a 10 dpi y en ninguna otra muestra tratada con FU. Sin embargo, tampoco podemos descartar que dicha mutación haya sido seleccionada en una de las tres réplicas como consecuencia de la selección positiva por la presencia del 5-FU. Sin embargo, dado que es una mutación en la proteína del movimiento es bastante improbable que sea una mutación de resistencia al 5-FU. No obstante, recientemente se ha encontrado en FMDV una mutación en la proteína no estructural 2C que puede modular el tipo de mutaciones que introduce la polimerasa viral en presencia de la ribavirina confiriéndole un

mecanismo de escape a la mutagénesis letal (Agudo *et al.*, aceptado para su publicación). Dada la dificultad de distinguir entre la selección positiva y negativa (Domingo *et al.*, 2012) habría que recurrir a estudios de fitness de dicho genotipo con la mutación G5626A en ausencia y presencia de 5-FU para averiguar si dicha mutación está relacionada con la resistencia al análogo.

Los espectros de mutantes en los que se detectaron reducciones de la infectividad de TMV se caracterizaron por no presentar mayor complejidad genética, medida como frecuencia de mutación, ni la heterogeneidad (índice de Shannon) que las poblaciones control. Tampoco se observaron diferencias en el número de mutaciones por molécula de TMV entre las muestras tratadas con 5-FU y control. Sin embargo al analizar en detalle los dos ORFs de la región secuenciada, RdRp y MP, se encontraron diferencias en la complejidad (estimada como diversidad genética media) de los espectros de mutantes. Así, durante el tratamiento con 5-FU la diversidad genética en la región RdRp fue menor que en la región MP y esta similar al control. Esas diferencias entre regiones desaparecieron después de un régimen de 21 días sin análogo. La menor diversidad de la RdRp durante el tratamiento sugiere una mayor restricción al cambio en esta región. Sin embargo, no hubo diferencias en el tipo de mutaciones entre ambas regiones. Las mutaciones predominantes en las poblaciones de TMV mutagenizadas con reducción en infectividad fueron transiciones A→G y U→C, las esperadas debido a la acción de FU ya que reemplaza residuos de uracilo en el RNA (Gleason and Fraenkel-Conrat, 1976; Hull, 2013). Esta abundancia de mutaciones asociadas al 5-FU no se encontró en las muestras a 5 dpi, en las que no hubo reducción significativa de infectividad vírica. Nuestros resultados concuerdan con los obtenidos de la Fase IIa de un ensayo clínico realizado en pacientes infectados con HIV con el análogo KP-1461 (una prodroga del análogo de desoxicitidina KP-1212) (Mullins *et al.*, 2011). En ese caso tampoco se detectó reducción de la carga viral ni aumento de la cantidad total de mutaciones durante el tratamiento de 124 días. Sin embargo, sí se detectaron variaciones en los espectros de mutantes de individuos tratados, con aumentos relativos de las mutaciones A→G y G→A, y algo menores de T→C and C→T, que son las predichas por el mecanismo de acción del análogo KP-1212. En ese estudio analizaron una secuencia de unos mil nucleótidos de la región *gag* del genoma de HIV por lo que no pudieron observar las diferencias en la

complejidad y así como en la distribución de las mutaciones entre zonas diferentes del genoma.

Cabe destacar que algunas de las cuasiespecies de las muestras control muestran una complejidad mayor que las muestras tratadas con 5-FU lo cual en principio era algo inesperado. Además a lo largo de la infección dichas cuasiespecies fueron aumentando su infectividad específica. La ganancia en *fitness* de una cuasiespecie en un ambiente determinado es el resultado de la optimización del conjunto de mutantes que mediante la competición entre sus componentes, tiende hacia un equilibrio mutación-selección. Esa ganancia en *fitness* en las poblaciones control puede deberse al origen clonal del virus y a su adaptación al ambiente *in vitro*. Nosotros observamos que la presencia del 5-FU alteró dicho equilibrio especialmente a la región RdRp, normalmente más restringida al cambio cuyo espectro de mutantes que mostró una mayor exploración hacia posiciones no permitidas a pesar de tener menor complejidad que el control. Estudios previos de LCMV en cultivos celulares (Grande-Perez et al., 2002) habían apuntado la importancia de la localización de las mutaciones en el genomas del virus durante la mutagénesis del virus ya que la pérdida de infectividad no se relacionaba directamente con el aumento de la frecuencia de mutación. Nuestros resultados indican que la región RdRp se ve más influida por la presencia del 5-FU y por tanto la hace idónea para el estudio de una terapia basada en el uso de defectores interferentes, siendo las primeras candidatas a ser testadas *in vivo* las mutaciones encontradas en esta región en presencia de 100 µg/ml de 5-FU.

En resumen, este es el primer trabajo que estudia la mutagénesis letal *in vivo* de un virus de RNA de plantas. La disminución de la infectividad de TMV debida al tratamiento con 5-FU presenta i) disminución de la infectividad específica, ii) no reducciones de la carga vírica iii) cambios en el espacio de secuencia explorado por los espectros de mutantes y iv) una secuencia consenso invariante y v) bajo grado de mutación de las moléculas genómicas de TMV, características que son el sello de identidad de la defección letal. Por tanto, nuestros resultados apoyan un modelo de defección letal *in vivo* para la mutagénesis letal y abre las puertas a emplear este sistema modelo para entender la cinética de acumulación de las moléculas defectoras y letales durante la defección letal, así como para profundizar en una estrategia antiviral basada en el uso de moléculas de virus defectoras de las regiones específicas del genoma más susceptibles a su acción

interferente. Además el sistema TMV-*N. tabacum* podría emplearse como una primera aproximación *in vivo* para entender las bases moleculares de la mutagénesis letal con otros agentes mutagénicos ya sea en monoterapia, en combinación o en administración secuencial con otros antivirales inhibidores no mutagénicos.

Anexos

TABLA A1. Mutaciones, cambio de codón, cambio de aminoácido y región de las cuasiespecies de TMV control y tratadas con 25, 50 y 100 µg/ml de 5-FU a 5,10 y 31dpi.

^a La numeración de los residuos del genoma de TMV se corresponde con la del genoma de *Tobacco mosaic virus* con número de acceso en Genbank AF273221

Muestra a 5 dpi	Mutación ^a	Cambio codón	Cambio aminoácido	Región
Control (1)	T4835C	GGU→GGC	Gly1589Gly	RdRp
	T4844C	GUU→GUC	Val1592Val	RdRp
	A4887T	AGA→UGA	Arg1607STOP	RdRp
	A4925G	AAA→AAG	Lys8Lys	MP
	T4942A	GUU→GAU	Val14Asp	MP
	A4956Del	AAA→AAA	Lys18Lys	MP
	G4963A	GGA→GAA	Gly21Glu	MP
	T5051C	AUU→AUC	Ile50Ile	MP
	T5104C	CUG→CCG	Cys68Pro	MP
	C5126T	CAC→CAU	His75His	MP
	C5143T	GCC→GUC	Ala81Val	MP
	A5247G	AAG→GAG	Lys115Glu	MP
	C5259T	CAA→UAA	Gln119STOP	MP
	T5398C	UUA→UCA	Leu166Ser	MP
	A5427G	AGA→GGA	Arg175Gly	MP
	A5446G	GAG→GGG	Glu182Gly	MP
	A5449G	AGA→GGA	Arg183Gly	MP
	A5528G	AAA→AAG	Lys209Lys	MP
	T5536A	AUC→AAC	Ile212Asn	MP
	G5635C	AGG→ACG	Arg244Thr	MP
Control (2)	T4441C	AUU→ACU	Ile1456Thr	RdRp
	C4492T	GCC→GUC	Ala1473Val	RdRp
	C4515T	CUG→UUG	Leu1481Leu	RdRp
	C4542T	CCG→UCG	Pro1490Ser	RdRp
	A4770G	AAU→GAU	Asn1566Asp	RdRp
	C4788T	CAG→UAG	Gln1572STOP	RdRp
	A4789T	CAG→CUG	Gln1572Leu	RdRp
	A4848G	AAA→GAA	Lys1592Glu	RdRp
	T4882Del	CUU→Del	Leu1603Del	RdRp
	T4884C	UUU→CUU	Phe1604Leu	RdRp
	A4923G	GGA→GGG	Gly4Gly	MP
	C5048T	UCA→UUA	Ser46Leu	MP
	G5166T	GUG→GUU	Val85Val	MP
	T5264C	GUC→GCC	Val118Ala	MP
	T5300C	AUG→ACG	Met130Thr	MP
	T5321Del	UUA→Del	Leu137Del	MP

Muestra a 5 dpi	Mutación ^a	Cambio codón	Cambio aminoácido	Región
	T5379C	UUU→UUC	Phe156Phe	MP
	T5393C	AUU→ACU	Ile161Thr	MP
	A5401T	AGA→UGA	Arg164STOP	MP
	T5483A	GUU→GAU	Val191Asp	MP
	T5508C	CCU→CCC	Pro199Pro	MP
	G5571Del	GGG→Del	Gly220Del	MP
	A5575Del	AUU→Del	Ile222Del	MP
	G5601A	CCG→CCA	Pro230Pro	MP
	A5656G	AAU→GAU	Asn249Asp	MP
	A5657Del	AAU→Del	Asn249Del	MP
	A5662G	AUC→GUC	Ile251Val	MP
	C5664A	AUC→AUA	Ile251Ile	MP
	C5664T	AUC→AUU	Ile251Ile	MP
FU25 (1)	G4530T	GGU→UGU	Gly1488Cys	RdRp
	C4542A	CCG→ACG	Pro1492Thr	RdRp
	A4588T	AAA→AUA	Lys1507Ile	RdRp
	T4865C	UAU→UAC	Tyr1599Tyr	RdRp
	A5029G	AAU→AGU	Asn43Ser	MP
	G5161T	GGU→UGU	Gly87Cys	MP
	G5162Del	GGU→GUG	Gly87Val	MP
	T5163C	GGU→GGC	Gly87Gly	MP
	A5243T	AAG→AUG	Lys114Met	MP
	T5257G	UUC→UGC	Phe119Cys	MP
	A5662G	AUC→GUC	Ile254Val	MP
FU50 (1)	C4385T	AUC→AUU	Ile1439Ile	RdRp
	A4596Ins	AAA→AAA	Lys1510Lys	RdRp
	G4812A	GUU→AUU	Val1582Ile	RdRp
	G4820T	AAG→AAU	Lys1584Asn	RdRp
	A5306del	AAC→AAG	Asn135Lys	MP
	T5561C	GUC→GCC	Val220Ala	MP
	A5661G	UUA→UUG	Leu253Leu	MP
	C5664A	AUC→AUA	Ile254Ile	MP
FU100 (1)	G4501T	GGU→GUU	Gly1476Val	RdRp
	A4555C	CAC→CCC	His1494Pro	RdRp
	C4671T	CUA→UUA	Leu1533Leu	RdRp
	T4866C	UUG→CUG	Leu1600Leu	RdRp
	T5359G	UGU→GGU	Cys153Gly	MP
	G5360A	UGU→UAU	Cys153Tyr	MP
	C5664T	AUC→AUU	Ile254Ile	MP
FU100 (2)	T4494G	UUU→GUU	Phe1476Val	RdRp

Muestra a 5 dpi	Mutación ^a	Cambio codón	Cambio aminoácido	Región
FU100 (3)	G4676A	AAG→AAA	Lys1536Lys	RdRp
	G5338A	GUG→AUG	Val146Met	MP
	G5338T	GUG→UUG	Val146Leu	MP
	T5530C	UUU→CUU	Phe210Leu	MP
	C5664T	AUC→AUU	Ile254Ile	MP
	A4428G	AAC→GAC	Asn1454Asp	RdRp
	C4456T	GCC→GUC	Ala1463Val	RdRp
	G4544T	CCG→CCU	Pro1492Pro	RdRp
	A4576T	AAU→AUU	Asn1503Ile	RdRp
	T4754C	GUU→GUC	Val1562Val	RdRp
	A4960Del	AUG→Del	Met20	MP
	T5019C	GUU→GUC	Val39Val	MP
	C5148T	GAC→GAU	Asp82Asp	MP
	T5406C	AAU→AAC	Asn186Asn	MP
	A5581G	AGU→GGU	Ser227Gly	MP
	C5599T	CCG→UCG	Pro233Ser	MP
	A5662T	AUC→UUC	Ile254Phe	MP
Muestra a 10 dpi	Mutación ^a	Cambio codón	Cambio aminoácido	Región
Control (1)	T4577C	AAU→AAC	Asn1503Asn	RdRp
	T4593G	UUU→GUU	Phe1509Val	RdRp
	T4748C	UGU→UGC	Cys1560Cys	RdRp
	C4881T	CUU→UUU	Leu1605Phe	RdRp
	T5022C	GAU→GAC	Asp40Asp	MP
	C5100T	UAC→UAU	Tyr66Tyr	MP
	G5122A	GUC→AUC	Val74Ile	MP
	G5162T	GGU→GUU	Gly87Val	MP
	G5626A	GAU→AAU	Asp242Asn	MP
FU25 (1)	G4710A	GAU→AAU	Asp1548Asn	RdRp
	A4960del	AUG→Del	Met20	MP
	T5053C	UCA→CCA	Ser51Pro	MP
	G5122A	GUC→AUC	Val74Ile	MP
	G5313A	UGG→UGA	Trp137STOP	MP
	C5599T	CCG→UCG	Pro233Ser	MP
	T5613C	UAU→UAC	Tyr237Tyr	MP
	A5656G	AAU→GAU	Asn252Asp	MP
FU50 (1)	A4589G	AAA→AAG	Lys1507Lys	RdRp
	G4888T	AGA→AUA	Arg1607Ile	RdRp
	T5355C	GGU→GGC	Gly151Gly	MP
	T5037C	CAU→CAC	His45His	MP
	G5200C	GCC→CCC	Ala100Pro	MP
	T5251C	UUU→CUU	Phe117Leu	MP

Muestra a 5 dpi	Mutación ^a	Cambio codón	Cambio aminoácido	Región
	G5293T	GAC→UAC	Asp131Tyr	MP
	C5365T	CUU→UUU	Leu155Phe	MP
	A5552del	AAG→Del	Lys217	MP
	C5664A	AUC→AUA	Ile254Ile	MP
	C5664T	AUC→AUU	Ile254Ile	MP
FU100 (1)	A4528G	AAG→AGG	Lys1487Arg	RdRp
	T4883C	CUU→CUC	Leu1605Leu	RdRp
	T4896C	UUU→CUU	Phe1610Leu	RdRp
	T4974ins	UUA→UUU	Leu24Phe	MP
	G4983C	AUG→AUC	Met27Ile	MP
	A5109G	UUA→UUG	Leu69Leu	MP
	A5132Del	GAG→GGU	Glu77Gly	MP
	G5179A	GUG→AUG	Val93Met	MP
	A5207G	GAG→GGG	Glu102Gly	MP
	C5562A	GUC→GUA	Val220Val	MP
	G5626A	GAU→AAU	Asp242Asn	MP
	G5633A	GGA→GAA	Gly244Glu	MP
	G5635A	GGA→AGA	Gly245Arg	MP
	T5644C	UUU→CUU	Phe248Leu	MP
FU100 (2)	T5037C	CAU→CAC	His45His	MP
	T5223C	UCU→UCC	Ser107Ser	MP
	T5502C	GAU→GAC	Asp200Asp	MP
FU100 (3)	T5444G	GUG→GGG	Val181Gly	MP
	C5562T	GUC→GUU	Val220Val	MP
	T5593C	UUA→CUA	Leu231Leu	MP
	T5631C	UUU→UUC	Phe243Phe	MP
	A5662G	AUC→GUC	Ile254Val	MP
	C5664T	AUC→AUU	Ile254Ile	MP
Muestra a 31 dpi	Mutación ^a	Cambio codón	Cambio aminoácido	Región
Control (1)	T4442C	AUU→AUC	Ile1458Ile	RdRp
	T4502C	GGU→GGC	Gly1478Gly	RdRp
	A4956Del	ACA→ACA	Thr18Thr	MP
	T5022C	GAU→GAC	Asp40Asp	MP
	C5295G	GAC→GAG	Asp131Glu	MP
FU100 (1)	T4386C	UGG→CGG	Trp1440Arg	RdRp
	A4519G	UAC→UGC	Tyr1484Cys	RdRp
	C4975Del	CCG→CGU	Pro25Arg	MP
	A5045G	GAG→GGG	Glu48Gly	MP
	T5094C	AGU→AGC	Ser64Ser	MP
	T5280C	GCU→GCC	Ala126Ala	MP

Muestra a 5 dpi	Mutación ^a	Cambio codón	Cambio aminoácido	Región
	G5333T	AGA→AUA	Arg144Ile	MP
	A5440G	AAC→GAC	Asn180Asp	MP
	T5644C	UUU→CUU	Phe248Leu	MP
	A5661C	UUA→UUC	Leu253Phe	MP

Tabla A2. Tipos de mutaciones encontradas en los espectros de mutantes de TMV a 5 dpi

Tipos de mutaciones	CONTROL ^a			FU100 ^a		
	RdRp	MP	TOTAL	RdRp	MP	TOTAL
Transiciones	%	%	%	%	%	%
A→G	15,4	25,0	22,4	9,1	7,1	8,0
G→A	0,0	5,6	4,1	9,1	14,3	12,0
C→U	30,8	13,9	18,4	18,2	28,6	24,0
U→C	30,8	22,2	24,5	18,2	28,6	24,0
Total	76,9	66,7	69,4	54,5	78,6	68,0
Transversiones						
A→C	0,0	0,0	0,0	9,1	0,0	4,0
C→A	0,0	2,8	2,0	0,0	0,0	0,0
A→U	15,4	2,8	6,1	9,1	7,1	8,0
U→A	0,0	8,3	6,1	0,0	0,0	0,0
C→G	0,0	0,0	0,0	0,0	0,0	0,0
G→C	0,0	2,8	2,0	0,0	0,0	0,0
G→U	0,0	2,8	2,0	18,2	7,1	12,0
U→G	0,0	0,0	0,0	9,1	0,0	4,0
Total	15,4	19,4	18,4	45,5	14,3	28,0
Substituciones	92,3	86,1	87,8	100,0	92,9	96,0
Inserciones	0,0	0,0	0,0	0,0	0,0	0,0
Deleciones	7,7	13,9	12,2	0,0	7,1	4,0
Total	100,0	100,0	100,0	100,0	100,0	100,0
No sinónimas						
Cambio de sentido	53,8	50	51,0	54,5	57,1	56
Sin sentido	15,3	5,5	8,1	0	0	0
Cambio marco de lectura	7,6	13,8	12,2	0	7,1	4
Sinónimas	23	30,5	28,5	45,4	35,7	40
Total	100	100	100	100	100	100

^aMutaciones de los espectros de mutantes de las réplicas 1 y 2 control (88961 nt secuenciados) y réplicas 1, 2 y 3 de 100 µg/ml de 5-FU (82881 nt secuenciados) descritas en la **A1 del Anexo**

CAPÍTULO 2

QuasiFlow: Un novedoso flujo de trabajo bioinformático para el análisis integral de los espectros de mutantes de cuasiespecies virales.

Introducción

La secuenciación del DNA ha evolucionado considerablemente durante los últimos 50 años. La primera secuenciación de una molécula de RNA_t en 1965 (Holley et al., 1965) supuso un primer hito que condujo al desarrollo de metodologías de secuenciación de DNA cada vez más robustas, creándose en 1977 la secuenciación de “terminación de cadena” (*chain-terminating*) por Frederick Sanger (Sanger et al., 1977). Sanger et al. determinaron la primera secuencia completa del genoma de un organismo, la del bacteriófago ϕ (Sanger et al., 1977), que fue seguido por la secuenciación de multitud de organismos, incluyendo virus de DNA y RNA, como *Epstein-Barr virus* (EBV) en 1984 (Baer et al., 1984) y el virus de la inmunodeficiencia humana (HIV) en 1985 (Ratner et al., 1985). En la actualidad, han sido depositadas en GenBank más de 2,7 millones de secuencias de genómicas virales. Esta información ha sido fundamental para comprender la estructura de los genomas virales, su biología, la evolución, la diversidad, la transmisión, la patogénesis, así como la evasión de la respuesta inmune del huésped, los medicamentos antivirales y vacunas. A pesar del impresionante éxito de la secuenciación “Sanger” y el desarrollo de instrumentos de secuenciación de DNA automatizados a mediados de la década de 1980, estos métodos de secuenciación de primera generación no son de alto rendimiento, su escalabilidad es limitada y no son rentables para la secuenciación de un gran número de muestras o de genomas de gran tamaño. Es por ello que era necesario el desarrollo de una segunda generación de tecnologías de secuenciación que solventara alguna de estas limitaciones.

Tras el desarrollo de varios métodos novedosos para la secuenciación de DNA a finales de 1990 y principios de 2000 (Brenner et al., 2000; Ronaghi et al., 1996), se han incorporado al mercado diferentes tecnologías para componer lo que ha sido denominada como secuenciación de próxima generación (NGS), también conocido como “segunda generación”, “secuenciación masiva” o “secuenciación profunda”. Estas tecnologías de secuenciación masiva son capaces de generar entre tres y cuatro órdenes de magnitud más de información que la anterior secuenciación de Sanger y son significativamente menos costosas, teniendo en cuenta el coste por nucleótido

secuenciado (Liu et al., 2012; Loman et al., 2012). Durante los últimos 12 años se han desarrollado varios sistemas de secuenciación masiva, cada uno de ellos con sus propios parámetros de rendimiento propios de ese sistema, como el número de lecturas obtenidas, longitud de las lecturas, la precisión, el tiempo por *run*, coste, etc. (Liu et al., 2012; Quail et al., 2012). En 2004, *454 Life Sciences* (Branford, CT) presentó el primer instrumento basado en pirosecuenciación (Margulies et al., 2005), antes de ser adquirida por Roche (Basilea, Suiza). A este equipo le siguió la aparición del *Genome Analyzer* (GA, Solexa, Chesterford, Reino Unido) en 2005 (Bennett, 2004), ahora renombrado como Illumina (San Diego, CA). En 2007, se crea el primer sistema de secuenciación *SOLiD* por Applied Biosystems (Foster City, CA) (Valouev et al., 2008), mientras que el secuenciador Helicos (Bowers et al., 2009) y el *Ion Torrent Personal Genome Machine* (Rothberg et al., 2011), fueron comercializados por Life Technologies (Carlsbad, CA) en 2009 y 2011, respectivamente. *Pacific Biosciences* (Menlo Park, CA) introdujo el secuenciador en tiempo real de una sola molécula (*single-molecule sequencer*) en 2011 (Eid et al., 2009) y, más recientemente, *Oxford Nanopore Technologies* (Oxford, Reino Unido) dio a conocer el secuenciador de moléculas simples ultra-largas (Howorka et al., 2001). Todas estas metodologías han seguido desarrollándose con nuevos productos químicos e instrumentos más potentes, lo que resulta en niveles cada vez más altos de rendimiento de secuenciación y un constante abaratamiento de los costes. Como resultado, el uso de la secuenciación masiva continúa extendiendo su uso tanto en la investigación y la práctica clínica, y el campo de la virología no es una excepción (Barba et al., 2014; Barzon et al., 2013; Radford et al., 2012).

1. Plataformas actuales de secuenciación masiva

Actualmente el mercado de la secuenciación masiva está copado por cuatro plataformas: 454TM (454 Life Sciences/Roche, Branford, CT) (Margulies et al., 2005), Illumina[®] (Illumina, Inc. de San Diego, CA) (Bennett, 2004), Ion (Ion Torrent TorrentTM/Life Technologies, Sur San Francisco, CA) (Rothberg et al., 2011), y PacBio[®] (Pacific Biosciences, Menlo Park, CA) (Eid et al., 2009). En general, las cuatro tecnologías son capaces de generar una información muy valiosa a nivel de secuenciación (Archer et al., 2012); sin embargo, existen diferencias fundamentales y significativas entre la

cantidad y la calidad de los datos, y esto hace que cada plataforma sea más o menos útil en los diferentes campos de la investigación, dependiendo de la aplicación que se le vaya a dar a dicha secuenciación.

La elección de una u otra plataforma de NGS puede llegar a ser determinante a la hora de obtener los resultados deseados. Aunque hasta la fecha la mayoría de los estudios publicados en virología han utilizado sistemas 454TM o Illumina®, tal vez debido al hecho de que estas eran las dos primeras metodologías disponibles, todas las tecnologías de NGS siguen mejorando y se están utilizando en multitud de estudios virológicos (Barba et al., 2014; Barzon et al., 2013; Gibson et al., 2014; Radford et al., 2012).

Los factores más importantes en la selección de una plataforma de secuenciación incluyen el tamaño de lectura o el tamaño del genoma en estudio, su complejidad (incluyendo contenidos en G+C), y la profundidad de la cobertura y la precisión necesaria. Para la secuenciación *de novo* del genoma, lo más apropiado sería contar con una longitud de lectura larga. Para la secuenciación por amplicones, la plataforma Roche 454 e Illumina Miseq son las más adecuadas debido a la longitud de sus lecturas, aunque 454 tiene la desventaja de ser poco competitiva en cuanto al precio con respecto a Illumina. Para RNA-seq y proyectos que requieran *coverage* alto, Illumina y SOLiD pueden ofrecer el mejor rendimiento y coste (Radford et al., 2012). Roche 454 tiene la longitud más larga de lectura, Illumina HiSeq 2500 cuenta con la mayor generación de lecturas y el coste más económico, y Solid 5500 xlw tiene la tasa de error más baja (Liu et al., 2012). Según (Barba et al., 2014) las plataformas de secuenciación de Illumina son las más utilizadas en los proyectos de virología de plantas actualmente.

2. Aplicaciones de la secuenciación masiva en Virología

La NGS ha sido la gran impulsora de todos los estudios genómicos realizados recientemente y la virología ha sido capaz de adaptarse a ella tomando una particular ventaja sobre el uso de esta metodología. En primer lugar, ahora es más sencillo y asequible que nunca secuenciar genomas virales completos. Del mismo modo, la identificación y clasificación de nuevos virus y de los ya conocidos, caracterizando poblaciones virales sin la necesidad de cultivarlos o clonarlos previamente a la secuenciación, la epidemiología molecular, la diversidad viral y la evolución, la transmisión y la patogénesis, y la virología médica, se han beneficiado enormemente de

la utilización de la NGS. El número de publicaciones mediante secuenciación masiva relacionadas con la virología se ha disparado desde 2008, en particular los relacionados con los virus humanos como el HIV (Quiñones-Mateu et al., 2014). El abaratamiento de los costes ha permitido que el estudio de variantes virales, incluyendo las poblaciones virales complejas, mediante secuenciación masiva se haya perfilado como una herramienta ideal para un amplio número de aplicaciones en los estudios de poblaciones virales.

3. La secuenciación masiva en el estudio de la diversidad viral

Hasta ahora, incluso con la llegada de las NGS, es muy común encontrar estudios de cuasiespecies basados en secuenciación Sanger, como se ha mostrado en el Capítulo 1 de la presente Tesis. Este tipo de análisis para el estudio de cuasiespecies virales necesita una gran mano de obra además de la producción de una gran multitud de clones moleculares. A día de hoy la NGS se está imponiendo a la aproximación por Sanger puesto que son capaces de generar hasta 10^7 moléculas para una sola muestra. Esto las hace ideales para estudiar la diversidad intra e inter-hospedador, su dinámica de adaptación e incluso el estudio de la progresión de la enfermedad. Por ejemplo, para el caso del HIV se han descrito estudios por NGS que han secuenciado el genoma completo, mientras que otros se han centrado en el estudio de determinadas regiones genómicas (Archer et al., 2012; Fisher et al., 2012; Hedskog et al., 2010; Lataillade et al., 2012; Mild et al., 2011; Mitsuya et al., 2008; Mortier et al., 2013; Recordon-Pinson et al., 2013; Rozera et al., 2009; Shao et al., 2013; Varghese et al., 2009). Otros estudios se han centrado en el análisis de la evolución intra-paciente del virus, tales como la caracterización de la transmisión del HIV y la persistencia de variantes minoritarias (Yin et al., 2012), la estimación de los tiempos de infección (Poon et al., 2011), la descripción de la dinámica de la evolución intra-huésped durante el curso de la infección con o sin tratamiento antirretroviral (Armenia et al., 2012; Hedskog et al., 2010; Henn et al., 2012; Recordon-Pinson et al., 2013; Rozera et al., 2009; Tsibris et al., 2009). Otros autores han analizado patrones de hipermutación (Knoepfel et al., 2011), la evolución viral en diferentes partes del huésped (Lee et al., 2012) o la respuesta del virus frente al sistema inmune del huésped (Wu et al., 2011; Xiao et al., 2013; Zhang et al., 2013), y el estudio del *fitness* de diferentes variantes resistentes de drogas (Brumme et al., 2013). En el

caso de HCV, la secuenciación masiva ha permitido el estudio de variantes minoritarias virales y la diversidad de las cuasiespecies intra-hospedador (Gregori et al., 2014; Ninomiya et al., 2012), la identificación de variantes virales de transmisión (Bull et al., 2011; Escobar-Gutiérrez et al., 2012), y los factores asociados con la patogénesis del HCV. Además, los estudios basados en NGS permiten la identificación de un número significativo de RNA no codificantes (RNAnc) asociados al carcinoma hepático causado por la infección con el HCV (Hou and Bonkovsky, 2013).

Los datos de NGS han confirmado que los virus de DNA también pueden generar mezclas complejas de genotipos virales en sujetos infectados y que la selección positiva podría haber contribuido a la divergencia de las diferentes cepas. Este es el caso de citomegalovirus humano (HCMV), que establece infecciones latentes de por vida en los seres humanos pudiendo reactivarse y causar una enfermedad grave, potencialmente mortal en pacientes inmunocomprometidos. Un estudio demostró una alta variabilidad intra-hospedador de HCMV en receptores de trasplante de pulmones infectados con este virus (Görzer et al., 2010) y en los recién nacidos con infección congénita de HCMV (Renzette et al., 2011).

Aunque existen gran cantidad de estudios de diversidad poblacional en virus humanos, el campo de la virología de plantas aún no ha explotado la NGS como método de estudio de cuasiespecies, a pesar de que la secuenciación masiva ofrece una mejor alternativa para diseccionar las poblaciones genéticamente complejas de virus de plantas. Por ejemplo, en una publicación pionera en el estudio de cuasiespecies de virus de plantas a través de NGS (Simmons et al., 2012), la secuenciación masiva de 5 regiones del genoma del *Zucchini Yellow Mosaic Virus* en el hospedador *Cucurbita pepo*, se llevó a cabo bajo dos condiciones de inoculación diferentes: vía vector infectivo y vía inoculación mecánica. Los resultados sugirieron que el vector o los cuellos de botella impuestos por el huésped (asociados respectivamente con la transmisión y la infección sistémica) son menos estrictos de lo que se suponía anteriormente. Estos resultados también mostraron que algunas mutaciones se fijaron en las poblaciones de áfidos, pero se mantuvieron en baja frecuencia en las plantas inoculadas mecánicamente.

Los pocos estudios realizados hasta ahora en virus de plantas emplearon la mayoría de las veces amplicones de los virus (Fabre et al., 2012; Martínez et al., 2012; Montarry et

al., 2011; Morroni et al., 2013), a pesar de que diferentes partes del genoma viral puede estar sometido a distintas presiones selectivas (Tugume et al., 2010; Yu et al., 2010) a excepción de estudios como el realizado con el genoma completo del *Potato virus Y* (Kutnjak et al., 2015) en el que hacen un estudio pormenorizado de la diversidad poblacional de este virus en el hospedador, o el estudio realizado con el geminivirus *Mungbean Yellow Mosaic India Virus* (Yadav et al., 2015).

Estos escasos estudios ponen en evidencia la necesidad de seguir no sólo mejorando las técnicas de secuenciación sino también la optimización de las herramientas y software informático que va a utilizarse para el análisis de las secuencias.

4. Software bioinformático para análisis de datos procedentes de NGS

Las funciones de software para el análisis de datos de secuenciación masiva se pueden clasificar en cuatro categorías generales que incluyen I) la alineación de las lecturas, II) *base-calling* o detección de polimorfismos, III) la exploración de la reconstrucción del genoma y su posterior anotación y IV) la síntesis *de novo*. Existen multitud de paquetes de software que han sido desarrollados para cada una de estas cuatro categorías. Por ejemplo, existe una gran variedad de herramientas bioinformáticas para la reconstrucción *de novo* de genomas a partir de secuencias disponibles (Paszkiwicz and Studholme, 2010; Zhang et al., 2011).

Muchas compañías de secuenciación comerciales, como Roche, Illumina, SOLiD, etc., ofrecen sus servicios de software a los clientes para diferentes tipos de análisis de secuenciación de DNA o RNA.

Hacer una revisión completa del software disponible para el de control de calidad y análisis cuantitativo NGS está más allá del alcance de esta Introducción, pero han sido objeto de varios artículos de revisión, (Horner et al., 2010; Miller et al., 2010; Pabinger et al., 2014; Pevzner and Shamir, 2011). Estas herramientas bioinformáticas y algoritmos se están desarrollando constantemente y mejorándose para estar al día.

El mayor problema al que se enfrentan estas herramientas bioinformáticas a la hora de estudiar la diversidad de las poblaciones virales son los falsos positivos originados por la amplificación por PCR y por secuenciación, ya que estos pueden introducir errores propios, lo que podría ser mal interpretado como mutaciones o polimorfismos. Es por ello que los estudios de los datos de NGS tienen que desarrollar protocolos y algoritmos

para estimar el error experimental, y para filtrar los resultados positivos de los falsos. Por lo tanto, se hace necesario el desarrollo de herramientas específicas para análisis de variantes de baja frecuencia y para el estudio de poblaciones virales.

De hecho, la determinación de un umbral de detección adecuado es un aspecto clave en todas estas herramientas de diagnóstico. Aun es objeto de debate si debido al error intrínseco de la técnica puede considerarse suficiente la detección de una sola lectura viral para valorarla como parte de la variabilidad genética de una muestra analizada por NGS. Teniendo en cuenta que las estrategias actuales NGS implican una o más etapas de PCR, es fácil entender que los enfoques basados en NGS serán tan susceptibles a los problemas de error en las muestras. En muchos casos, estos problemas con los errores son sólo importantes en secuenciaciones con bajo número de lecturas.

Otra cuestión se refiere a la adaptación de los protocolos y técnicas para las especies particulares de plantas, o tejidos de estas, debido a que es conocido que algunas especies de plantas contienen sustancias que interfieren con la extracción de ácidos nucleicos o con la manipulación enzimática posterior (transcripción inversa, amplificación por PCR, RCA).

Estos nuevos enfoques basados en NGS tienen un enorme potencial en el campo del estudio de cuasiespecies virales. Este potencial ya está ampliamente ilustrado por la continua aparición de nuevas herramientas para el estudio de diferentes aspectos de las poblaciones virales descritas en la literatura científica, pero tienen el hándicap de la dificultad de uso de las herramientas actuales disponibles. Si se desea generalizar el uso de la secuenciación masiva en el estudio de cuasiespecies existe una clara necesidad de desarrollar herramientas bioinformáticas simples, fiables y fáciles de usar que permitan la identificación rápida y sencilla de las variantes virales presentes en una muestra. Aunque se ha avanzado mucho en esa dirección, todavía hay mucho margen de mejora y para el desarrollo de programas y algoritmos más rápidos y más eficientes (Massart et al., 2014). Esto se hace particularmente importante en el caso de intentar indexar las lecturas procedentes de NGS para lograr secuencias virales completas o haplotipos dentro de poblaciones altamente divergentes, ya que esta tecnología aún está en pleno desarrollo.

En este Capítulo II hemos analizado las carencias actuales en el estudio de las cuasiespecies de virus de plantas a través de secuenciación masiva para desarrollar un software que permita el estudio de variantes minoritarias en el genoma secuenciado llamado QuasiFlow, que sea de uso sencillo, y que profundice en la comparación entre muestras secuenciadas para sacarle el máximo rendimiento a la nueva generación de secuenciaciones de las que disponemos hoy en día. Además, demostramos la versatilidad de QuasiFlow analizando muestras de plantas coinfectadas con geminivirus y profundizamos en el estudio de las heteroplasmas mitocondriales con el uso del nuevo software.

Materiales y Métodos

1. Material biológico.

1.1. Plantas

Para la comprobación del funcionamiento de QuasiFlow se utilizó material biológico procedente del IHSM “La Mayora”, amablemente cedido por el Dr. Enrique Moriones, que consistió en 2 muestras de hojas de *Solanum lycopersicum* infectadas con los geminivirus TYLCV-MId, TYLCV-IL y TYLCMaV. Una de las muestras fue recogida 15 días después de la agroinfección (tomate 15 dpi) y la otra muestra a los 30 días (tomate 30 dpi).

2. Extracción y purificación de ácidos nucleicos.

2.1. Extracción de DNA de *S. lycopersicum* infectadas con TYLCV

Para la extracción de DNA, las dos muestras de *S. lycopersicum* se pulverizaron en un mortero con Nitrógeno líquido y se almacenaron a -80°C hasta su uso. A 150 mg de tejido pulverizado se añadieron 400 µl de tampón de extracción (200 mM Tris- HCl, pH 7,5, 200 mM NaCl, 25 mM EDTA, 0,5%, 0,1% β-mercaptoetanol), se agitó con vortex 2 minutos y se centrifugó a 12000 g durante 5 min a temperatura ambiente. El sobrenadante se transfirió a un nuevo tubo al que se le añadieron 300 µl de fenol:cloroformo 1:1 (v/v) equilibrado a pH 8, mezclando bien las fases antes de usarlo. A continuación se centrifugó a 12000 g durante 5 min a temperatura ambiente y la fase acuosa fue depositada en un nuevo tubo. El DNA se precipitó con 300 µl de isopropanol tras incubación durante 2 min a temperatura ambiente. Tras centrifugar a 12000 g durante 5 min el precipitado se lavó una vez con etanol al 100% y otra con etanol al 70%. Por último, el DNA se secó a temperatura ambiente, se resuspendió en 100 µl de tampón TE (10 mM Tris HCl pH 8,1 mM EDTA) y se guardó a 4°C.

3. Tratamiento enzimático de ácidos nucleicos

3.1. Tratamiento con endonucleasas de restricción

Las enzimas de restricción empleadas fueron suministradas por la casa comercial TaKaRa. Los tratamientos del DNA se llevaron a cabo a la temperatura recomendada durante al menos 1 h utilizando 1-2 unidades de enzima por microgramo de DNA. Las

enzimas de restricción se inactivaron según las instrucciones del fabricante, generalmente incubándolas durante 30 min a 65°C.

Tratamiento de las extracciones de DNA total con *DpnI*

Las muestras de DNA procedentes de plantas de *S. lycopersicum* agroinfectadas se trataron con la endonucleasa *DpnI* para eliminar los restos de DNA viral procedente de *A. tumefaciens* de la muestra y no el DNA viral replicado en el interior de la planta infectada. Esta enzima de restricción reconoce la secuencia GATC en la que las adeninas estén metiladas. Dado que la metilación en eucariotas tiene lugar en las citosinas, *DpnI* digiere el DNA procedente de organismos procariotas pero no el procedente de organismos eucariotas.

3.2. Eliminación de endonucleasas de las muestras.

Para la purificación del DNA tratado con endonucleasas se utilizó el kit comercial StrataClean Resin (Agilent Technologies), siguiendo las instrucciones indicadas por el fabricante. El funcionamiento de este kit se basa en una purificación libre de fenoles en la cual, la fase sólida de resina contiene grupos hidroxilos que reaccionan con las proteínas con la misma afinidad que lo harían los grupos hidroxilos del fenol, permitiendo una separación del DNA y las endonucleasas. Para ello, en todos los casos, se añadió un volumen de resina del doble de la muestra, mezclando con vórtex durante 15 segundos. La mezcla se incubó 1 minuto a temperatura ambiente y se centrifugó a 2000 g durante otro minuto. El sobrenadante que contiene el DNA fue transferido a un nuevo tubo y se repitió el proceso desde el principio una vez más, obteniendo el mismo volumen inicial de la muestra.

4. Reacciones de amplificación de ácidos nucleicos

4.1. Amplificación por círculo rodante (RCA)

Para amplificar el genoma viral presente en los extractos de DNA total, 1 µl del mismo a una concentración de 130 ng/µl, se utilizó para amplificar por círculo rodante (*Rolling circle amplification*, RCA) usando el kit Illustra™ ThempliPhi (GE Healthcare) siguiendo las indicaciones del fabricante (ver apartado 4.1. de Materiales y Métodos del Capítulo I).

Para reducir la acumulación de errores aleatorios que pueda introducir la DNA polimerasa $\Phi 29$ cada muestra se amplificó por duplicado uniéndose el producto de ambas amplificaciones en un volumen final de 20 μ l. Dado que los errores generados en ambas reacciones no serán idénticos conseguimos que el ratio mutaciones reales/mutaciones RCA se incremente al doble.

5. Electroforesis de ácidos nucleicos

5.1. Electroforesis en geles de agarosa

La separación e identificación de moléculas de DNA se realizó mediante electroforesis horizontal en geles de agarosa (Agarose SPI, Duchefa) disuelta en tampón TBE (Tris-Base 89 mM, EDTA 2 mM y Ácido Bórico 89 mM). La concentración de agarosa varió en función del tamaño de los fragmentos a separar, oscilando entre 0.8-1% (p/v). A la solución de agarosa fundida se añadió Sybr Safe (1 μ l/ml) para la posterior visualización de los ácidos nucleicos. Cada muestra se mezcló con tampón de carga para geles de agarosa en proporción 5:1 (v/v). La electroforesis se realizó en cubetas horizontales con tampón TBE, a un voltaje constante de 3-4 V/cm. Los geles se fotografiaron con un documentador de geles ChemiDoc XRS+ (Biorad) empleando el programa Image Lab®. Como patrones de tamaño molecular se empleó 1 Kb Plus DNA Ladder de Invitrogen. La concentración de DNA se determinó midiendo la densidad óptica (DO) a una longitud de onda de 260 nm con un espectrofotómetro NanoVue Plus (GE Healthcare) o por cuantificación en geles de agarosa en comparación con patrones de DNA de cantidades conocidas.

6. Secuenciación

La secuenciación de las muestras de *S. lycopersicum* infectadas y amplificadas por RCA fue llevada a cabo en Macrogen, Seul, Corea del Sur. Estas muestras fueron secuenciadas mediante la tecnología Illumina MiSeq. Para ello las genotecas se crearon con el kit Truseq DNA PCR-free con un tamaño de inserto de 550 nt. Las muestras de DNA se suministraron a Macrogen a una concentración de 100 ng/ μ l en un volumen final de 20 μ l. La tecnología de Illumina MiSeq permitió obtener lecturas pareadas de una longitud de 301 nt cada una. Las 2 muestras de plantas de tomate infectadas con begomovirus

se secuenciaron en el mismo *run*, obteniendo un número de lecturas pareadas en bruto de $1,7 \times 10^6$ en el caso de la muestra a 15 dpi y $1,6 \times 10^6$ para la de 30 dpi.

Los datos de la secuenciación de mitocondrias humanas (empleados en el estudio de variabilidad mitocondrial presentado en este capítulo) proceden de la base de datos SRA del NCBI <http://www.ncbi.nlm.nih.gov/sra>, son propiedad de Sanger Institute y pertenecen al Proyecto PRJEB5005 con consentimiento de uso público. Todas son secuenciaciones Illumina Miseq pareadas a partir de librerías de 400 nt. La **TABLA 1** muestra los números de acceso de cada una de las 43 muestras de DNA mitocondrial secuenciadas por el Sanger Institute, procedentes de un número variable de hombres y mujeres. El DNA mitocondrial secuenciado fue extraído de sangre, línea celular linfoblastoide o saliva. Según descripción de los autores, estos datos pertenecen a un estudio comparativo entre el DNA mitocondrial humano y el DNA mitocondrial de gorilas, siendo el DNA mitocondrial humano el control de dicho estudio.

Sexo	BioSample	Experimento	Organismo	Run	ID SRA	Descripción muestra	ID Quasiflow
Hombre	SAMEA2258558	ERX534475	<i>Homo sapiens</i>	ERR576063	ERS372267	ADN de línea celular linfoblastoide	Male_Lymphoblastoid_1
Hombre	SAMEA2258548	ERX534450	<i>Homo sapiens</i>	ERR576038	ERS372257	ADN de línea celular linfoblastoide	Male_Lymphoblastoid_2
Hombre	SAMEA2258556	ERX534451	<i>Homo sapiens</i>	ERR576039	ERS372265	ADN de línea celular linfoblastoide	Male_Lymphoblastoid_3
Hombre	SAMEA2258549	ERX534462	<i>Homo sapiens</i>	ERR576050	ERS372258	ADN de línea celular linfoblastoide	Male_Lymphoblastoid_4
Hombre	SAMEA2258550	ERX534474	<i>Homo sapiens</i>	ERR576062	ERS372259	ADN de línea celular linfoblastoide	Male_Lymphoblastoid_5
Hombre	SAMEA2258559	ERX534487	<i>Homo sapiens</i>	ERR576075	ERS372268	ADN de línea celular linfoblastoide	Male_Lymphoblastoid_6
Hombre	SAMEA2258552	ERX534498	<i>Homo sapiens</i>	ERR576086	ERS372261	ADN de línea celular linfoblastoide	Male_Lymphoblastoid_7
Hombre	SAMEA2258508	ERX534445	<i>Homo sapiens</i>	ERR576033	ERS372217	ADN de saliva	Male_Saliva_1
Hombre	SAMEA2258512	ERX534493	<i>Homo sapiens</i>	ERR576081	ERS372221	ADN de saliva	Male_Saliva_2
Hombre	SAMEA2258509	ERX534457	<i>Homo sapiens</i>	ERR576045	ERS372218	ADN de saliva	Male_Saliva_3
Hombre	SAMEA2258510	ERX534469	<i>Homo sapiens</i>	ERR576057	ERS372219	ADN de saliva	Male_Saliva_4
Hombre	SAMEA2258513	ERX534505	<i>Homo sapiens</i>	ERR576093	ERS372222	ADN de saliva	Male_Saliva_5
Hombre	SAMEA2258514	ERX534517	<i>Homo sapiens</i>	ERR576105	ERS372223	ADN de saliva	Male_Saliva_6
Hombre	SAMEA2258537	ERX534508	<i>Homo sapiens</i>	ERR576096	ERS372246	ADN de sangre	Male_Blood_1
Hombre	SAMEA2258522	ERX534518	<i>Homo sapiens</i>	ERR576106	ERS372231	ADN de sangre	Male_Blood_2
Hombre	SAMEA2258542	ERX534473	<i>Homo sapiens</i>	ERR576061	ERS372251	ADN de sangre	Male_Blood_3
Hombre	SAMEA2258539	ERX534532	<i>Homo sapiens</i>	ERR576120	ERS372248	ADN de sangre	Male_Blood_4
Mujer	SAMEA2258557	ERX534463	<i>Homo sapiens</i>	ERR576051	ERS372266	ADN de línea celular linfoblastoide	Female_Lymphoblastoid_1
Mujer	SAMEA2258543	ERX534485	<i>Homo sapiens</i>	ERR576073	ERS372252	ADN de línea celular linfoblastoide	Female_Lymphoblastoid_2
Mujer	SAMEA2258551	ERX534486	<i>Homo sapiens</i>	ERR576074	ERS372260	ADN de línea celular linfoblastoide	Female_Lymphoblastoid_3
Mujer	SAMEA2258544	ERX534497	<i>Homo sapiens</i>	ERR576085	ERS372253	ADN de línea celular linfoblastoide	Female_Lymphoblastoid_4
Mujer	SAMEA2258545	ERX534509	<i>Homo sapiens</i>	ERR576097	ERS372254	ADN de línea celular linfoblastoide	Female_Lymphoblastoid_5
Mujer	SAMEA2258553	ERX534510	<i>Homo sapiens</i>	ERR576098	ERS372262	ADN de línea celular linfoblastoide	Female_Lymphoblastoid_6
Mujer	SAMEA2258546	ERX534521	<i>Homo sapiens</i>	ERR576109	ERS372255	ADN de línea celular linfoblastoide	Female_Lymphoblastoid_7
Mujer	SAMEA2258554	ERX534522	<i>Homo sapiens</i>	ERR576110	ERS372263	ADN de línea celular linfoblastoide	Female_Lymphoblastoid_8
Mujer	SAMEA2258547	ERX534533	<i>Homo sapiens</i>	ERR576121	ERS372256	ADN de línea celular linfoblastoide	Female_Lymphoblastoid_9
Mujer	SAMEA2258555	ERX534534	<i>Homo sapiens</i>	ERR576122	ERS372264	ADN de línea celular linfoblastoide	Female_Lymphoblastoid_10
Mujer	SAMEA2258516	ERX534446	<i>Homo sapiens</i>	ERR576034	ERS372225	ADN de sangre	Female_Blood_1

Mujer	SAMEA2258523	ERX534530	<i>Homo sapiens</i>	ERR576118	ERS372232	ADN de sangre	Female_Blood_2
Mujer	SAMEA2258524	ERX534447	<i>Homo sapiens</i>	ERR576035	ERS372233	ADN de sangre	Female_Blood_3
Mujer	SAMEA2258532	ERX534448	<i>Homo sapiens</i>	ERR576036	ERS372241	ADN de sangre	Female_Blood_4
Mujer	SAMEA2258517	ERX534458	<i>Homo sapiens</i>	ERR576046	ERS372226	ADN de sangre	Female_Blood_5
Mujer	SAMEA2258525	ERX534459	<i>Homo sapiens</i>	ERR576047	ERS372234	ADN de sangre	Female_Blood_6
Mujer	SAMEA2258533	ERX534460	<i>Homo sapiens</i>	ERR576048	ERS372242	ADN de sangre	Female_Blood_7
Mujer	SAMEA2258518	ERX534470	<i>Homo sapiens</i>	ERR576058	ERS372227	ADN de sangre	Female_Blood_8
Mujer	SAMEA2258526	ERX534471	<i>Homo sapiens</i>	ERR576059	ERS372235	ADN de sangre	Female_Blood_9
Mujer	SAMEA2258534	ERX534472	<i>Homo sapiens</i>	ERR576060	ERS372243	ADN de sangre	Female_Blood_10
Mujer	SAMEA2258511	ERX534481	<i>Homo sapiens</i>	ERR576069	ERS372220	ADN de sangre	Female_Blood_11
Mujer	SAMEA2258519	ERX534482	<i>Homo sapiens</i>	ERR576070	ERS372228	ADN de sangre	Female_Blood_12
Mujer	SAMEA2258527	ERX534483	<i>Homo sapiens</i>	ERR576071	ERS372236	ADN de sangre	Female_Blood_13
Mujer	SAMEA2258535	ERX534484	<i>Homo sapiens</i>	ERR576072	ERS372244	ADN de sangre	Female_Blood_14
Mujer	SAMEA2258520	ERX534494	<i>Homo sapiens</i>	ERR576082	ERS372229	ADN de sangre	Female_Blood_15
Mujer	SAMEA2258528	ERX534495	<i>Homo sapiens</i>	ERR576083	ERS372237	ADN de sangre	Female_Blood_16
Mujer	SAMEA2258521	ERX534506	<i>Homo sapiens</i>	ERR576094	ERS372230	ADN de sangre	Female_Blood_17
Mujer	SAMEA2258529	ERX534507	<i>Homo sapiens</i>	ERR576095	ERS372238	ADN de sangre	Female_Blood_18
Mujer	SAMEA2258530	ERX534519	<i>Homo sapiens</i>	ERR576107	ERS372239	ADN de sangre	Female_Blood_19
Mujer	SAMEA2258515	ERX534529	<i>Homo sapiens</i>	ERR576117	ERS372224	ADN de sangre	Female_Blood_20
Mujer	SAMEA2258531	ERX534531	<i>Homo sapiens</i>	ERR576119	ERS372240	ADN de sangre	Female_Blood_21
Mujer	SAMEA2258540	ERX534449	<i>Homo sapiens</i>	ERR576037	ERS372249	ADN de sangre	Female_Blood_22
Mujer	SAMEA2258541	ERX534461	<i>Homo sapiens</i>	ERR576049	ERS372250	ADN de sangre	Female_Blood_23
Mujer	SAMEA2258536	ERX534496	<i>Homo sapiens</i>	ERR576084	ERS372245	ADN de sangre	Female_Blood_24
Mujer	SAMEA2258538	ERX534520	<i>Homo sapiens</i>	ERR576108	ERS372247	ADN de sangre	Female_Blood_25

Tabla 1: Descripción de las muestras mitocondriales extraídas de SRA del NCBI pertenecientes al Sanger Institute.

7. Bioinformática

El desarrollo del software para el análisis bioinformático de cuasiespecies virales secuenciadas por NGS, QuasiFlow, fue desarrollado en colaboración con el Dr. M. Gonzalo Claros y el estudiante predoctoral Pedro Seoane Zonjic, del Centro de SuperComputación y Bioinnovación de la UMA (SCBI).

QuasiFlow es un flujo de trabajo que requiere de un conjunto específico de programas especialmente diseñados para secuencias nucleotídicas procedentes de secuenciación masiva. Estos programas están diseñados todos para Linux y funcionan bajo entorno UNIX. Para una correcta comprensión del capítulo se ha considerado necesario definir los conceptos algoritmo, programa, flujo de trabajo, variable y *script*.

Algoritmo: es una lista bien definida, ordenada y finita de operaciones que permite hallar la solución a un problema. Dado un estado inicial y una entrada, a través de pasos sucesivos y bien definidos se llega a un estado final, obteniendo una solución.

Programa: es un conjunto de instrucciones que una vez ejecutadas realizarán una o varias tareas en una computadora. Al conjunto general de programas, se lo denomina software y así se refiere al equipamiento o soporte lógico de una computadora digital.

Flujo de trabajo: es una tarea bioinformática diseñada para abarcar varios programas y/o algoritmos, que ejecutados en serie o en paralelo obtienen un resultado final.

Variable: en el campo de la bioinformática, hace referencia a un nombre simbólico o identificador que tiene asociado un valor almacenado con una cantidad de información conocida. Esta separación entre identificador y valor permite que el nombre simbólico sea usado independientemente de la información exacta que representa. De esta forma se puede modificar el valor de estas variables sin tener que modificar el algoritmo al que esté asociado dicha variable.

Script: es un programa normalmente muy simple, que por lo general se almacena en un archivo de texto plano. En esta tesis se utilizaron *scripts* para complementar algunos aspectos de QuasiFlow, bien aportando un *input* para el flujo, o bien analizando algún *output* de este, ejecutándose de manera independiente.

7.1. Equipo informático

Para el desarrollo de la herramienta bioinformática QuasiFlow se ha utilizado el siguiente equipo: MacBook Pro (Retina, 15 pulgadas, finales de 2013), con procesador 2

GHz Intel Core i7, 8 GB 1600 MHz DDR3 de memoria RAM y tarjeta gráfica Intel Iris Pro 1536 MB. El trabajo se realizó bajo el sistema operativo OS X El Capitan versión 10.11.2. Sin embargo, como QuasiFlow necesita de entornos mucho más potentes para llevar a cabo su función, el flujo se ejecutó en el supercomputador PICASSO perteneciente al SCBI de la Universidad de Málaga. PICASSO tiene un hardware compuesto por 2312 cores, 21 TB de memoria RAM, 750 TB de memoria ROM y 31 TFLOPS. Una información más detallada acerca de los recursos de PICASSO está disponible la página web <http://www.scbi.uma.es/site/>.

7.2. Programas

El hecho de que QuasiFlow solo pueda ser ejecutado en PICASSO obliga al usuario a tener aplicaciones en su terminal que permitan el acceso al supercomputador. Estas aplicaciones dependerán del sistema operativo nativo en el ordenador de cada usuario. En nuestro caso, al tratarse de un entorno IOS, se optó por el uso de los programas Cyberduck y iTerm2 con los que accedimos al servidor SFTP (SSH Secure File Transfer) de PICASSO.

- Entorno IOS

Cyberduck: es una aplicación cliente de código abierto de FTP y SFTP, WebDAV, Rackspace Cloud, Google Docs, and Amazon S3 para los sistemas operativos Mac OS. Cyberduck está escrito en lenguaje Java y usa la interfaz de usuario Cocoa. Su función principal es la transferencia segura de archivos entre una computadora local y otra remota. Además provee de funciones básicas de manejo de archivos (copia, creación de ficheros, lectura de archivos, etc). Incorpora un protocolo de seguridad SSH.

iTerm2: Es un cliente para el protocolo de red SSH. Este protocolo se utiliza para ejecutar una sesión remota en un computador, a través de una red. En términos simples, iTerm se ejecuta en IOS y se le indica que debe conectar a una máquina UNIX (PICASSO en nuestro caso). Cualquier comando que se escriba en la pantalla será enviado directamente a PICASSO, y la respuesta de este, será mostrada en la pantalla. Se puede trabajar en PICASSO como si se estuviera presente en su consola de comandos.

- Entorno UNIX

Todos los programas utilizados para la creación del flujo de trabajo QuasiFlow son de código libre y su funcionamiento está ampliamente explicado en sus respectivas

publicaciones. El uso que se le ha dado a cada uno de estos programas en el Flujo se ha descrito en el apartado 1. de Resultados. Todos los programas en el flujo de trabajo están basados en lenguajes de programación Ruby, Perl o R.

Autoflow: Es un gestor de flujos de trabajo que permite construir cualquier flujo de programas que trabajen en UNIX. En este trabajo fue utilizado como base para la construcción de QuasiFlow.

Seqtrim Next: Es un flujo de trabajo que mediante diferentes algoritmos permite el pre-procesado de las secuencias procedentes de una secuenciación masiva de cualquier tipo (Falgueras et al., 2010). QuasiFlow lo utiliza para filtrado de secuencias de baja calidad, búsqueda de secuencias similares a una referencia dada y eliminación de indeterminaciones en las secuencias y otros contaminantes.

BWA: Este software se empleó para el indexado de las secuencias de referencia y el posterior mapeo de las secuencias obtenidas por secuenciación masiva .

Samtools: Este software contiene varias utilidades para la manipulación de alineamientos en formato SAM. En el desarrollo de Quasiflow se utilizó para la selección, fusión e indexación de secuencias previamente alineadas.

Bowtie: Al igual que BWA, empleamos este programa para el indexado de la secuencia de referencia que posteriormente será usada por otros programas.

Samstats: Programa diseñado para obtener diferentes estadísticas a partir de un archivo SAM.

Vcftools: este programa ofrece métodos accesibles para trabajar con variaciones genéticas procedentes de secuenciaciones complejas como las de los virus. Necesita un archivo .vcf como entrada.

Varscan: Es un programa diseñado para el análisis de variantes dentro de los datos de secuenciación masiva. Se puede utilizar para detectar diferentes tipos de variación, en nuestro caso para la detección de SNPs e InDels.

Haploclique: Se trata de un software especializado en la reconstrucción de la estructura de cuasiespecies virales a partir de datos de secuenciación masiva. Permite reconstrucción de haplotipos y estimación de su abundancia.

R-Bioconductor: Es un software que proporciona herramientas para el análisis y comprensión de los datos obtenidos por otros programas de análisis genómico de NGS.

Clustal omega: Programa que permite el estudio filogenético de secuencias procedentes de la reconstrucción de los haplotipos de la cuasiespecie viral.

VirVarSeq: Este software ofrece la posibilidad de estudiar variaciones minoritarias en codones. Su estrategia de filtrado es adaptativa y para cada muestra analizada ofrece un umbral de calidad optimizado.

ViReMa: Este programa detecta y enumera cualquier evento de recombinación o eventos de fusión en genomas virales secuenciados por NGS.

Circos: Es un software para la visualización de datos en un plano circular que lo hace ideal para la representación de genomas y variables.

- **Scripts**

Para la optimización de Quasiflow se hizo necesaria la creación de Controles *in silico* a través de los cuales evaluar cada uno de los filtros y algoritmos aplicados al flujo, así como la calibración de las variables utilizadas para distinguir entre el error perteneciente a la muestra y el producido por la propia técnica de secuenciación. Los controles que se utilizaron están extensamente descritos en el apartado 1.1 de Resultados

La secuencia a partir de la cual se obtuvieron los controles *in silico* pertenece a TYLCV-Mld (Genbank: AF071228). Con el programa Art-Illumina se creó una secuenciación *in silico* con las características de una secuenciación Illumina MiSeq de esta secuencia de TYLCV-Mld. Las características de esta secuenciación fueron: lecturas de 250 nt de longitud pareadas secuenciadas a partir de librerías de 550 nt, *coverage* por nucleótido de 60.000X y error de secuenciación asociado a la técnica que el programa añade de forma automática. Esta secuenciación fue nuestro control negativo.

Para la elaboración del resto de controles fue necesaria la creación de dos *scripts*. Un primer *script* que introdujese de forma aleatoria un número deseado de mutaciones por cada secuencia del control negativo (nombrado como *mutator.rb*), y un segundo *script* que crease recombinaciones aleatorias en las secuencias de los controles anteriores (*recombinator.rb*). El código fuente de ambos puede consultarse en el Anexo *scripts*. De esta forma tendremos un control negativo sin recombinaciones y otro con recombinaciones, además de un control positivo sin recombinaciones y otro con recombinaciones, todos ellos partiendo de la misma secuenciación *in silico* realizada con ART-Illumina.

Para el análisis del número de variables en función de su frecuencia (análisis iterativo de variables) se creó un script llamado *iterative_varsScan.rb*, el cual a partir de una frecuencia mínima y máxima de aparición de variables y un número “n” de repeticiones dado, ejecuta el análisis “n” veces desde la frecuencia mínima a la frecuencia máxima de detección de variables.

7.3. Análisis estadísticos

Todos los análisis estadísticos mostrados en este capítulo han sido llevados a cabo de forma automática por QuasiFlow a partir del análisis comparativo entre muestras. Todos los análisis estadísticos se han escrito en R. Los análisis estadísticos llevados a cabo por QuasiFlow se componen de ANOVA para el análisis comparativo entre muestras o variables de muestras (origen de la muestra, dpi, genotipo, etc), índice de Shannon y diversidad nucleotídica calculados a partir de la reconstrucción de haplotipos y análisis de componentes principales (PCA) de todas las muestras analizadas y todas sus variables. Todos estos análisis estadísticos han sido descritos en el apartado 1.2.2. de Resultados.

Resultados

1. Análisis bioinformático de secuencias de virus mediante la herramienta QuasiFlow.

1.1. Descripción

Para examinar la variabilidad de las cuasiespecies virales a partir de datos de NGS, hemos desarrollado el programa QuasiFlow, una herramienta de código libre que, a modo de flujo de trabajo, utiliza software especializado en el análisis de cuasiespecies virales con datos de secuenciación masiva. QuasiFlow permite analizar tanto datos de secuencias de DNA como de RNA viral. Está orientado al estudio de secuencias pareadas Illumina HiSeq y MiSeq con independencia del tamaño de la genoteca. Esta restricción es debido a que actualmente la secuenciación basada en Illumina es la que ofrece la mayor calidad de secuencias y con la mayor longitud (Barba et al., 2014), lo cual es determinante a la hora de la reconstrucción de haplotipos.

QuasiFlow minimiza el tiempo de análisis de los datos con respecto al uso de los programas que lo forman de manera independiente y, por lo tanto, permite incrementar el número de muestras analizadas en paralelo, automatizando lo máximo posible este tipo de análisis. La consecuencia inmediata de esta automatización y análisis en paralelo conseguidos por QuasiFlow es hacer reproducible cualquier estudio de análisis de variantes.

El programa QuasiFlow se ejecuta desde una terminal y se encuentra instalado en el Supercomputador PICASSO perteneciente al SCBI de la Universidad de Málaga, estando a disposición de los investigadores.

Para facilitar la comprensión del flujo de trabajo y su funcionamiento, en este apartado hemos evitado detallar aspectos sobre el código fuente del programa, que se encuentra disponible en el Anexo QUASIFLOW y en el manuscrito Díaz-Martínez *et al.* (en preparación).

1.2. Desarrollo

QuasiFlow se estructura en dos bloques independientes entre sí: un primer bloque se encarga del análisis individual de cada una de las muestras sometidas a estudio y el segundo bloque realiza un análisis estadístico comparativo de todas las muestras de

forma conjunta. El diagrama de la **Figura 1** muestra un resumen de las entradas y salidas de cada bloque del programa.

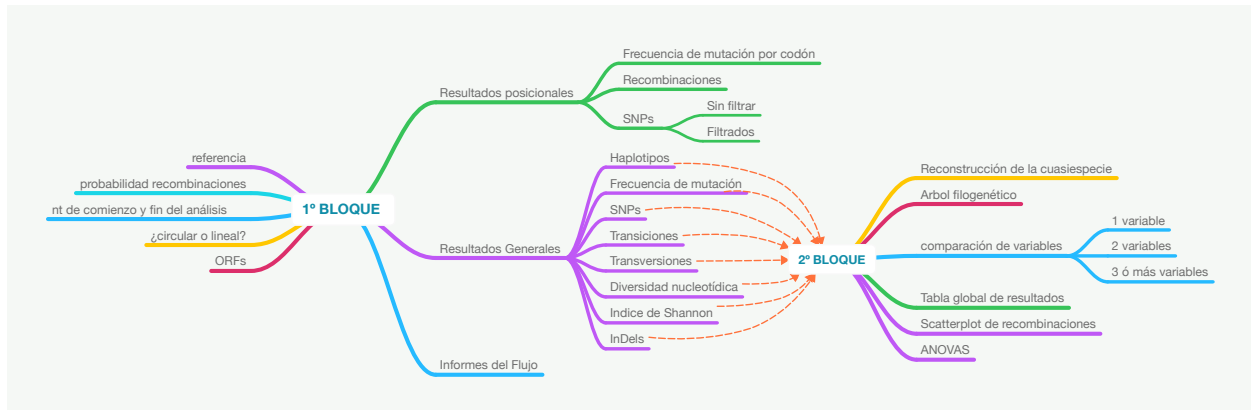


Figura 1. Representación esquemática del funcionamiento de la herramienta QuasiFlow. Se estructura en dos bloques: el primero requiere la introducción de variables específicas de cada muestra por parte del usuario. De este primer bloque se obtienen resultados posicionales y generales, además de informes sobre el funcionamiento del flujo de trabajo. A partir de los resultados generales obtenidos de todas las muestras analizadas, QuasiFlow ejecuta el segundo bloque, dando como resultado un análisis comparativo de las muestras.

- Análisis de muestras individuales (Bloque 1)

Este primer bloque de análisis requiere definir el número mínimo de variables que QuasiFlow utilizará para caracterizar las muestras. Estas variables se encuentran dentro de un archivo de texto plano (*Daemon*.sh*) que funciona a modo de archivo ejecutable (**Figura 2**), dentro del cual deben completarse los siguientes campos:

```
#!/usr/bin/env bash

virus_reference=$general_path"/virus.fasta"
raw_files=$general_path"/raw_reads"
consensus_start='1'
consensus_end='2791'
circular='-c'
rec_prob='0.00007'
orfs='1551116ldir;3151258ldir;1641359lrev;3211100lrev;11521135lrev;13001134lrev'
group1="2698:2536:2332:1879:1813:1513:1512:1283:1280:1242:1046:1045"
group2="931:923:872:860:839:821:760:731:707:572:539:501:470:355:163:141:100:88:55:46"

seqtrimnext_chunk_length=5000
```

Figura 2. Ejemplo del archivo ejecutable *Daemon*.sh* para muestras de TYLCV. Se muestra la configuración recomendable para el estudio de TYLCV-Mld con los campos (nombres en blanco) de las variables que se deben completar (verde) para ejecutar la herramienta QuasiFlow.

- *Virus reference*: nombre de la secuencia de referencia y directorio en el que se encuentra ubicada.
- *Raw files*: directorio donde se han depositado los archivos de lecturas.
- *Consensus start & Consensus end*: nucleótido de inicio y fin (con respecto a la secuencia de referencia) para reconstruir la secuencia consenso.
- *Circular*: variable que indica si el DNA sujeto de estudio es circular o lineal. Esta variable es importante a la hora de evaluar si las recombinaciones encontradas entre el último nucleótido de la secuencia y el primero son reales o se trata de una falsa recombinación producida por la linearización de la secuencia de referencia.
- *Rec prob*: probabilidad mínima de recombinación para que sea aceptada como una recombinación real (0,007% por defecto)
- *ORFs*: para determinar cada ORF es necesario definir el nucleótido de inicio de cada ORF, su longitud (en número de codones) y la situación de cada ORF en la cadena directa o reversa.
- *Group1 & Group2*: permite indicar si una posición nucleotídica que cambia con respecto a la referencia aparece en la reconstrucción de la cuasiespecie.
- *SeqTrimNext Chunk Length*: variable que define la procedencia de la secuenciación. Hay que indicar valor “20000” si las lecturas proceden de una secuenciación HiSeq o un valor “5000” (por defecto) si las lecturas proceden de MiSeq. Para comenzar su funcionamiento, Quasiflow requiere además una lista (*samples_to_process.lst*) con los nombres de todas las muestras que se deseen analizar en paralelo.

Esta primera fase de análisis individual de muestras se lleva a cabo en tres etapas: preparación de las muestras, análisis de las muestras y recopilación de resultados. La **Figura 3** muestra de forma resumida todos los procesos que lleva a cabo QuasiFlow de forma rutinaria durante esta primera fase.

- **Fase de preparación de muestras (Bloque 1)**

En esta etapa se lleva a cabo el indexado de la secuencia de referencia. A su vez, se produce el filtrado de secuencias a través del programa SeqtrimNext (Corchado et al., 2007; Falgueras et al., 2010), que realiza una limpieza de secuencias de vectores,

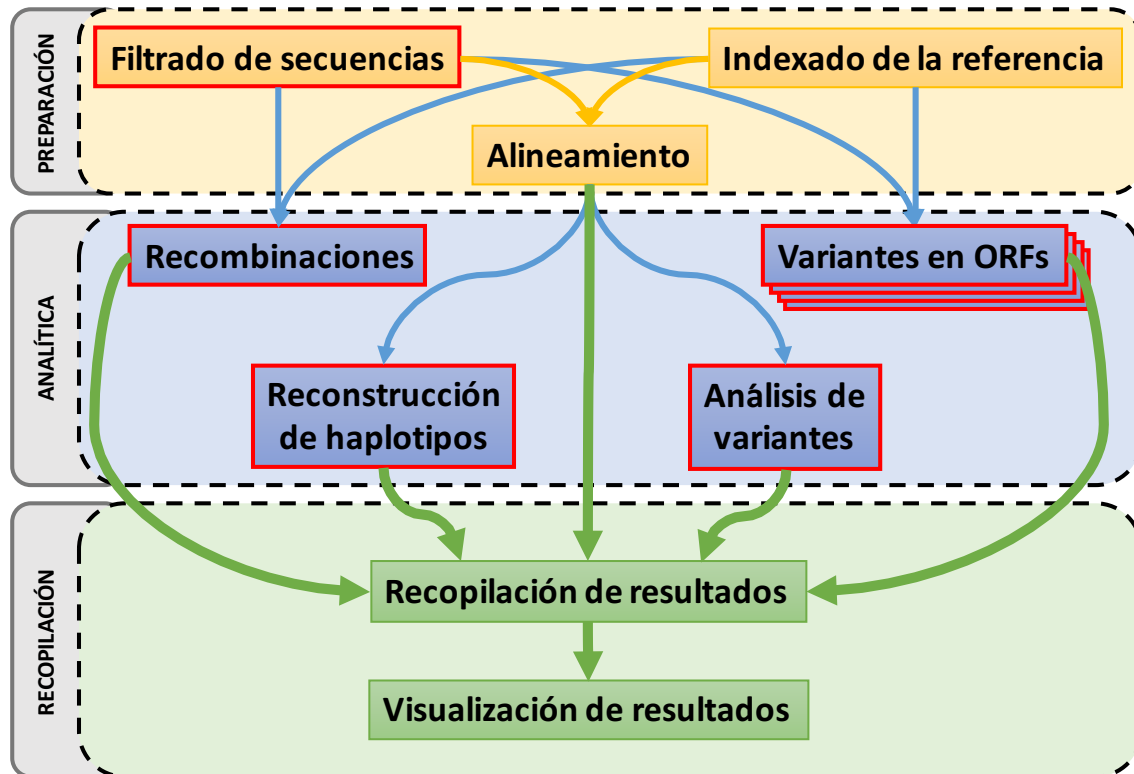


Figura 3. Fases del análisis de una muestra con QuasiFlow. La fase primera (recuadro amarillo) se caracteriza por el filtrado, indexado y alineamiento de las muestras. En la segunda fase (recuadro azul) se lleva a cabo el análisis de variantes como secuencias recombinantes, análisis de SNPs e InDels, variantes por ORF, reconstrucción de haplotipos, etc... La tercera fase (recuadro verde) recopila y procesa todos los datos obtenidos de las fases anteriores. Las fases del análisis con control de funcionamiento se han marcado en rojo.

adaptadores, secuencias poli-A y poli-T, secuencias N y también realiza un filtrado por tamaños de secuencias obtenidas (todas las secuencias de un tamaño inferior a 100 nt son eliminadas). *SeqtrimNext* incluye una opción para eliminar DNA contaminante, como por ejemplo lecturas que pertenezcan a DNA del hospedador, requiriendo este *script* la secuencia .fasta del DNA que se quiera identificar como contaminante. Dado que no siempre se dispone de la secuencia del hospedador de referencia, en QuasiFlow hemos adaptado este *script* para capturar las lecturas que muestran similitud de secuencia con la secuencia de referencia del virus. Las secuencias víricas filtradas de esta forma son las que entran en el flujo de trabajo, quedando el resto de secuencias retenidas en este paso. El filtro de calidad de secuencias que usa QuasiFlow es de un Phred>30 (QC>30), lo cual implica que las lecturas que pasen el filtro tienen una probabilidad de 1×10^{-3} errores por nucleótido secuenciado. Este error, aunque bajo, no

es suficiente para poder estudiar las variantes menos representadas en la cuasiespecie; es por ello que a lo largo del flujo, QuasiFlow aplica nuevos filtros para conseguir datos de secuencia con una menor probabilidad de error que se explicarán en el siguiente apartado de análisis de muestras individuales.

- **Fase de análisis de muestras individuales (Bloque 1)**

La fase analítica es el núcleo central del bloque de análisis de muestras individuales (Figura 3). Todos los análisis que se realizan en esta fase se ejecutan en paralelo, de forma que se consigue una reducción considerable en el tiempo empleado. Es en esta fase central donde se lleva a cabo el estudio de recombinaciones, reconstrucción de haplotipos, búsqueda de variantes nucleotídicas y donde se analiza cada uno de los ORFs para estudiar la variabilidad en los codones.

El análisis de recombinación se realiza a través del programa ViReMa (Routh and Johnson, 2014). Tal como se muestra en la **Figura 4**, este programa crea una secuencia “semilla” de una longitud de 20 nt en cada lectura secuenciada (**Figura 4A**) y la compara con la secuencia de referencia (**Figura 4B**). Una vez ha encontrado una identidad del 100%, se extiende la región alineada hasta que deja de coincidir la secuencia problema con la referencia en un nucleótido (**Figura 4C**), en cuyo caso salta a la siguiente posición (**Figura 4D**) y así hasta un total de 8 posiciones (**Figura 4E**). Si no encuentra identidad de secuencia, se crea una nueva semilla de 20 nt que se compara a lo largo de la secuencia de referencia y, en el caso de encontrar una coincidencia, anota el último nucleótido de la referencia que fue mapeado con la primera semilla y el primer nucleótido mapeado con la segunda semilla, interpretando que ha ocurrido una recombinación entre ambos puntos (**Figura 4F**).

Para la reconstrucción de haplotipos, QuasiFlow utiliza la herramienta HaploClique (Töpfer et al., 2014). Este programa crea “*superlecturas*” a partir de un alineamiento de lecturas de secuencias que compartan un 100% de identidad en el 90% de su longitud de secuencia. Cuando HaploClique encuentra dos secuencias en la misma zona de la referencia que solapan entre sí en un 90% de su longitud, pueden darse dos situaciones, o bien que la zona solapante entre ambas secuencias sean idénticas, o bien que no lo sean. En el primer caso, el programa interpreta que se trata de dos lecturas de un mismo haplotipo y las fusiona creando una *superlectura* que tendrá un tamaño ligeramente

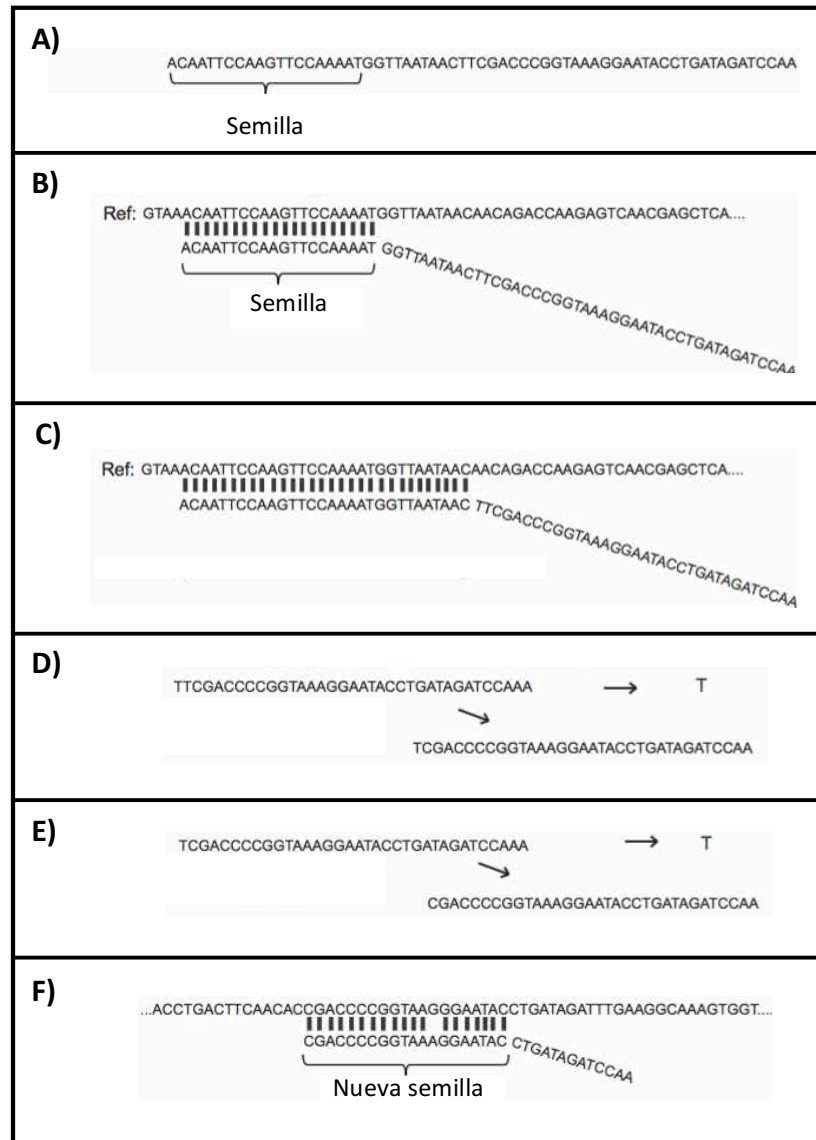


Figura 4. Análisis de secuencias recombinantes con el programa ViReMa. Se crea una secuencia “semilla” de una longitud de 20 nt en cada lectura secuenciada (A) y la compara con la secuencia de referencia (B). Una vez ha encontrado una identidad del 100%, se extiende la región alineada hasta que deja de coincidir la secuencia problema con la referencia en un nucleótido (C). La lectura se segmenta y se crea una nueva lectura (D). Si esta nueva lectura no mapea a continuación del fragmento anterior se elimina el primer nucleótido creando una nueva semilla, en cuyo caso salta a la siguiente posición. Se repite el proceso y así hasta un total de 8 posiciones consecutivas (E). La nueva semilla busca correspondencia con la referencia en otra parte del genoma. En el caso de encontrarla crea una recombinación entre el último nucleótido del primer alineamiento y el primer nucleótido de

mayor que las lecturas de las que procede. En el segundo caso, aunque sólo exista una diferencia de un nucleótido, no creará la *superlectura* ya que en esta situación se interpreta que ambas lecturas pertenecen a haplotipos diferentes. El problema surge

cuando estas *superlecturas* no llegan a tener la longitud total de la referencia y como consecuencia, la reconstrucción de los haplotipos queda incompleta. Para solucionarlo hemos desarrollado un *script* integrado como parte del flujo de QuasiFlow que complementa esta reconstrucción de haplotipos. Dicho *script* está basado en la ejecución cíclica de este módulo de reconstrucción de haplotipos utilizando como nueva entrada las *superlecturas* generadas en el ciclo anterior. De esta forma se consigue aumentar de forma continua la longitud de los haplotipos reconstruidos. Además, a medida que aumenta la longitud de los haplotipos, el *script* baja el umbral de longitud solapante para facilitar la fusión de las *superlecturas*.

Otro de los módulos que se ejecutan en esta fase analítica es el del análisis de variantes SNPs e InDels. QuasiFlow genera un nuevo filtro que permite distinguir entre variantes reales en las lecturas y los errores producidos durante la secuenciación. Este módulo se basa en un algoritmo que sólo acepta como válidas las variantes que se repitan al menos 10 veces por cada 10000 nucleótidos secuenciados en cada posición de la referencia, es decir el 0,1% de un *coverage* mínimo de 10000X. Este algoritmo permite disminuir el error de secuenciación 30 veces (10 veces por cada nucleótido A, T, C o G, diferente para cada posición de la referencia, estimando que no hay sesgo en la introducción del nucleótido erróneo durante la secuenciación), lo que deja el umbral de error de secuenciación en $3,3 \times 10^{-5}$ errores por nucleótido analizado, alcanzando valores de fiabilidad muy cercanos a la secuenciación Sanger (Ewing et al., 1998; Noguchi et al., 2006). Los resultados de este módulo son complementados por el análisis de variaciones a nivel de codón para cada uno de los ORFs del virus. A través de ambos, es posible apreciar en las proteínas la repercusión de los cambios producidos en la secuencia de DNA.

- **Fase de recopilación de datos de muestras individuales (Bloque 1)**

Esta fase se encarga de recolectar todos los datos obtenidos a lo largo de la fase anterior generando una representación gráfica de las variantes más importantes, para dar una visión global de la estructura de la cuasiespecie. En este bloque se encuentran los *scripts* para calcular el índice de Shannon normalizado y la diversidad genética a partir de las reconstrucciones de las cuasiespecies (ver Anexo *Scripts*). También genera un directorio que contiene la salida en forma de Tablas de todos los análisis realizados:

- Tabla resumen: contiene datos generales de la muestra, número de lecturas, recombinaciones, número de SNPs e InDels, tipos de mutaciones, número de transiciones y transversiones, ratio Ts/Tv, frecuencia de mutación, índice de Shannon, diversidad nucleotídica, frecuencia de mutación de cada ORF por nucleótido, etc.
- Tabla SNPs: muestra la posición de todos los SNPs encontrados en la muestra, nucleótido de la referencia, nucleótido de cambio y frecuencia del cambio con respecto al *coverage*.
- Tabla InDels: muestra la posición de todos los InDels encontrados en la muestra, nucleótido de la referencia de cada InDel, secuencia del InDel, y frecuencia con respecto al *coverage*.
- Tabla *coverage*: muestra el *coverage* por nucleótido de la muestra analizada.
- Tabla de recombinaciones: muestra todas las recombinaciones encontradas en la muestra, posiciones que se encuentran en contacto en la recombinación, número de veces que se ha encontrado dicho evento y hebras del DNA que están implicadas en la recombinación.
- Tabla reconstrucción de haplotipos: muestra el número de lecturas creadas en cada ciclo y la longitud máxima alcanzada por los haplotipos reconstruidos.
- **Puntos de control (Bloque 1)**

A lo largo de todo el proceso de análisis de muestras, QuasiFlow genera un informe de funcionamiento en las fases más importantes del flujo, de forma que permite la trazabilidad de las secuencias analizadas y la comprobación del buen funcionamiento del programa en cada uno de los módulos. En la **Figura 3** se pueden observar, marcados en rojo, dónde se encuentran situados dichos puntos de control dentro de la fase de análisis de muestras individuales. De esta forma, todos los módulos que se ejecutan en la fase analítica crean archivos que contienen la ID de cada secuencia utilizada para cada uno de los procesos para que, si se desea, pueda hacerse un estudio manual extrayendo dichas secuencias. Además, durante el filtrado de secuencias que realiza SeqtrimNext se crea un archivo *.html* como el que muestra la **Figura 5**, indicando los resultados del filtrado por calidad y por longitud de lectura de las secuencias.

Para comprobar que QuasiFlow funciona de forma correcta y comprobar este primer bloque de análisis hemos creado *in silico* cuatro controles. Utilizando la herramienta

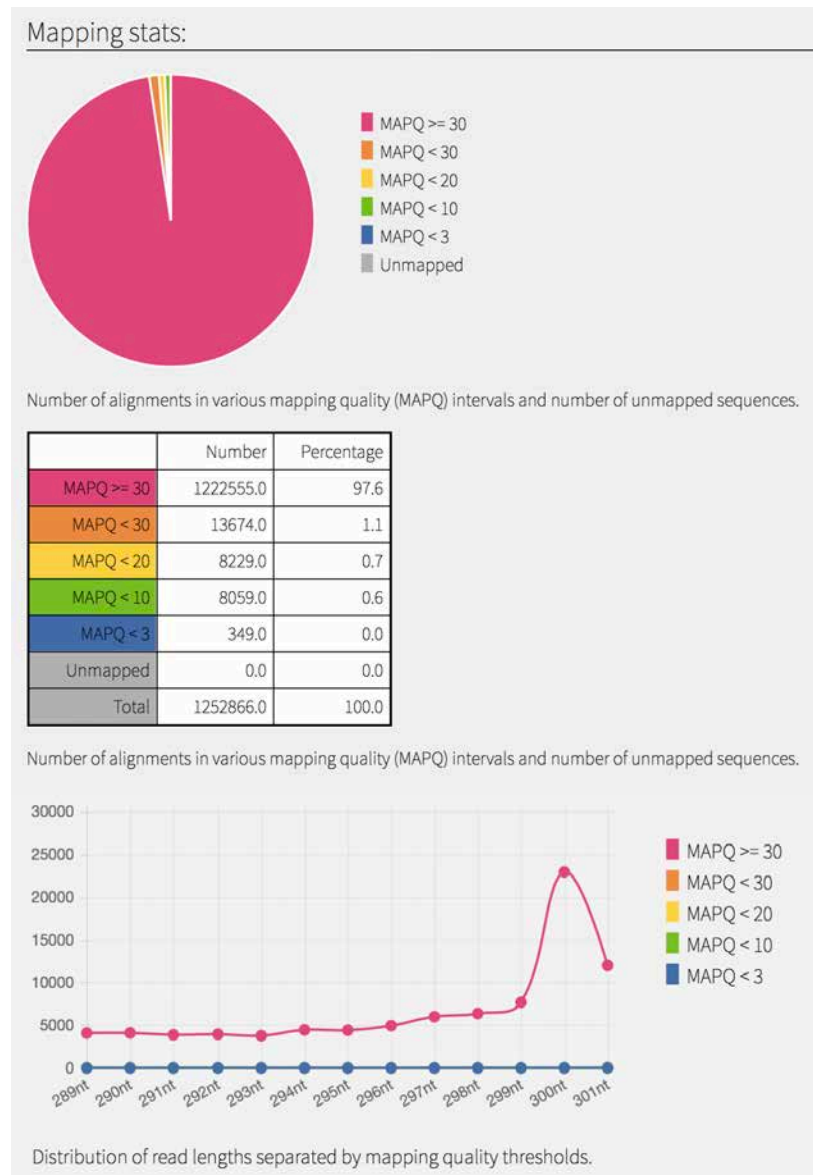


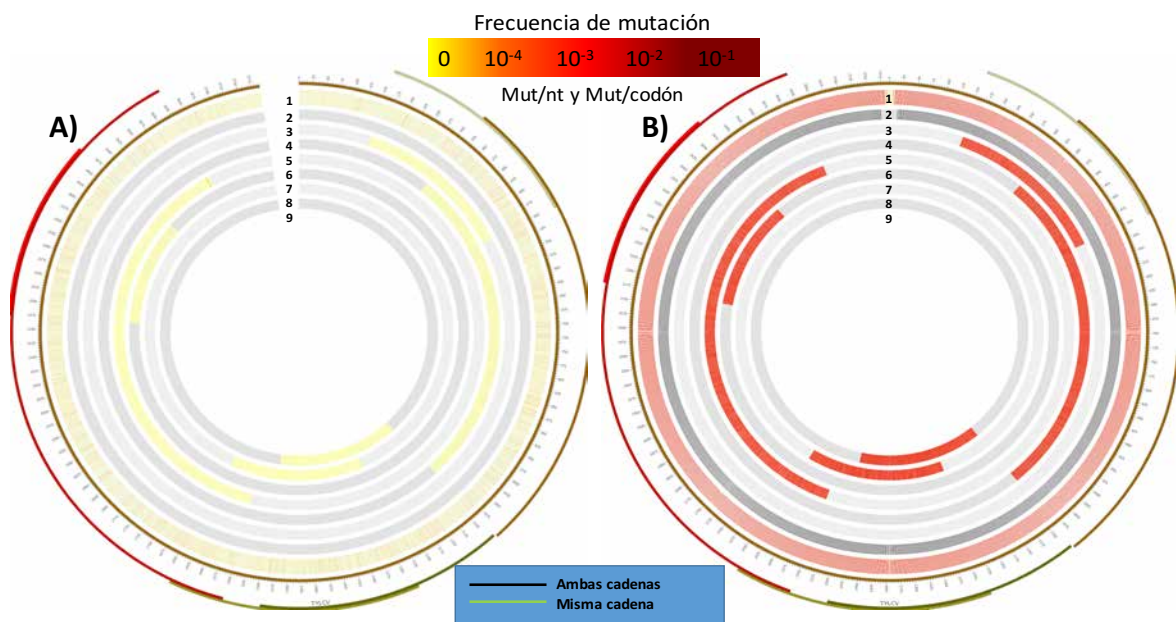
Figura 5. Representación gráfica generada a partir de los resultados obtenidos con SeqTrimNext tras la aplicación de filtros a las secuencias en estudio. En primer lugar se muestra un gráfico con los porcentajes de las calidades de cada lectura que va a entrar en el flujo de trabajo, separándolas en lecturas de calidad MAPQ(mapping quality)<30 (azul); <10 (verde); <20 (amarillo); <30 (naranja) y >=30 (rosa). Por último se indica mediante una gráfica la distribución de secuencias por su longitud (eje X), numero de secuencias (eje Y) y calidad (color).

ART-Illumina creamos una secuenciación masiva del virus TYLCV-MId con un porcentaje de errores en la secuencia similar a los producidos por la técnica de secuenciación Illumina MiSeq. La secuenciación creada a partir de este programa se caracteriza por tener un *coverage* medio de 60000X, con lecturas de 250 nt de longitud pareadas a partir de genotecas de 400 nt. Esta muestra serviría de control negativo de mutaciones. A

partir de esta secuenciación masiva creada *in silico* se generó otra secuenciación idéntica a la anterior pero con una acumulación de errores muy superior (hasta 6 errores por secuencia) a partir del *script* Mutator.rb de propia creación. Esta muestra hizo las veces de control positivo de mutaciones.

Tanto el control positivo como el control negativo sirvieron de molde para crear dos controles más con las mismas secuencias que los anteriores pero en los que un 60% de las secuencias incluían recombinaciones aleatorias (control positivo y negativo recombinaciones respectivamente). Esto se consiguió mediante la ejecución del *script* complementario al flujo llamado Recombinator.rb, de propia creación. De esta forma se pudo evaluar el comportamiento del flujo con muestras antagónicas.

Ambos controles negativo y positivo sin recombinaciones mostraron los resultados esperados. En caso del control negativo (**Figura 6A**), todos los errores de secuenciación



fueron filtrados y la representación gráfica de la muestra mostró tanto ausencia de variantes por nucleótido y por codones, como de recombinaciones. Por el contrario, el control positivo de mutaciones sin recombinaciones sí mostró cambios a lo largo de toda la secuencia tanto en nucleótidos como en codones, ambos con una alta frecuencia de mutación (**Figura 6B**).

En cuanto al control con recombinaciones sin mutaciones (**Figura 7A**) se observó cómo el alto grado de recombinación influye sobre los codones de los diferentes ORFs con una distribución de la frecuencia de mutación aleatoria dentro de cada ORF. Las recombinaciones no parecen afectar a la detección de variantes de nucleótido, ya que QuasiFlow consigue filtrarlas. En cuanto al control con recombinaciones sin mutaciones (**Figura 7B**), el flujo trató a todas las recombinaciones como artefactos por lo que fueron filtradas y eliminadas del análisis. Concluimos que QuasiFlow es una herramienta apta

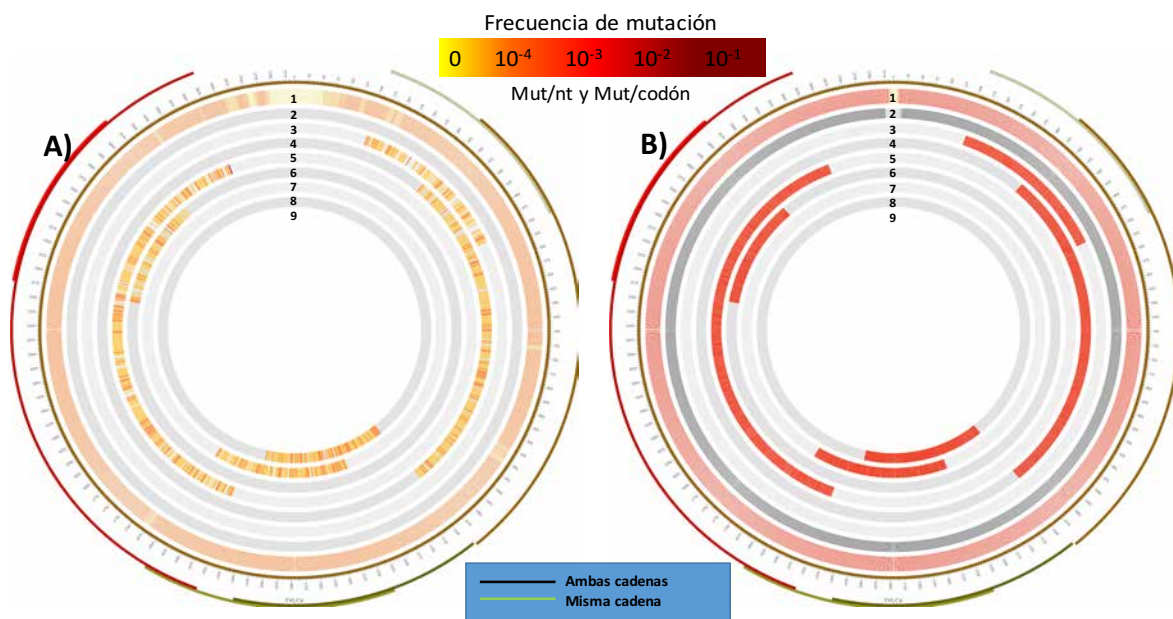


Figura 7. Representación gráfica del perfil de mutaciones de los controles de TYLCV-M1d *In silico* negativo sin recombinaciones (A) y positivo sin recombinaciones (B). La frecuencia de mutación de nucleótidos y de codones está expresada por un código de colores del amarillo (frecuencia de mutación no detectable por Quasiflow) al rojo (frecuencias de mutación cercanas a 10^{-1} mutaciones por nucleótido secuenciado en esa posición). Los círculos concéntricos del gráfico numeradas desde el exterior al interior representan: (1) Frecuencia de mutación en cada nucleótido de las lecturas previo al pase del filtro de calidad. (2) Posiciones mutadas dentro de la referencia que han pasado el filtro de calidad (líneas negras). Frecuencias de mutación de codones a nivel de triplete de bases para cada ORF determinado para este ensayo (3-8). Representación de puntos de recombinación, diferenciando entre recombinaciones entre cadenas de la misma polaridad (línea verde) o de polaridades distintas (línea negra), (9)

para el estudio de recombinación, ya que puede distinguir entre los eventos de recombinación reales de la muestra y los eventos artefactuales, dando validez a los resultados que arroja el flujo.

- **Análisis estadístico comparativo (Bloque 2)**

Al igual que ocurre en el primer bloque de análisis de muestras individuales de QuasiFlow, esta fase del flujo necesita que las muestras sean caracterizadas en un archivo de texto plano. En este archivo (llamado en el flujo *target.txt*) además del nombre de la muestra es necesario indicar qué parámetros se desean estudiar, y ubicar cada muestra dentro de dichos parámetros, elaborando una tabla como la Tabla 2.

Tabla 2. Ejemplo del archivo de texto plano *target.txt* con indicación de muestras a analizar junto con las variables que se van a comparar: tiempo y tipo de hospedador.

Muestra	Réplica	Tiempo	Hospedador
Col0_28_1	1	28d	Wildtype
Col0_28_2	2	28d	Wildtype
Tomato_30_1	1	28d	Tomato_1
Tomato_30_2	2	28d	Tomato_1

Una vez caracterizados todos los parámetros, QuasiFlow analiza todas las muestras mediante diferentes estrategias. Todos los análisis y estadísticas se han escrito en R, cuyo código puede verse en el Anexo FASE 2.

- ANOVA: QuasiFlow genera de forma automática una tabla donde indica si alguna de las variables estudiadas en las muestras individuales presentan un valor de $P < 0,05$.

- Análisis de una sola variable: en este apartado, el flujo crea gráficas donde representa la media de las réplicas biológicas con su desviación estándar frente a cualquiera de las variables en estudio (frecuencia de mutación, número de haplotipos, índice de Shannon, SNPs, etc.).

- Análisis de dos variables: en esta fase, el flujo crea gráficas donde se representan las muestras como puntos distribuidos en un entorno bidimensional donde cada uno de los ejes X e Y representan las variables a estudiar. QuasiFlow genera un recta patrón a partir de dichos puntos e indica si existe o no relación entre ambas variables.

- Análisis de componentes principales: este análisis es el más potente de todos los que realiza QuasiFlow. Genera un análisis integrado del comportamiento de las muestras al

estudiar todas las variables de forma conjunta. En este análisis, cada una de las variables se transforma en un vector con longitud y dirección únicas dentro de un espacio bidimensional. Esta transformación de las variables se aplica a cada una de las muestras a estudio, de forma que el sumatorio de todas las variables genera una posición única para cada muestra, agrupándose las muestras que sean similares y separando las muestras que difieran en una o más variables.

- Scatterplot de recombinaciones: para una mejor comprensión de patrones de recombinación, QuasiFlow permite agrupar los eventos de recombinación de las réplicas biológicas en un solo gráfico. De esta forma pueden identificarse las zonas más propensas a recombinar y los patrones de recombinación pueden ser comparados con otras muestras.

Todos los análisis que QuasiFlow realiza en su segunda etapa permiten normalizar los resultados de las variables que definen una cuasiespecie y estudiar las poblaciones virales a partir de réplicas biológicas dando una mayor robustez a cualquier diseño experimental que se desee desarrollar.

2. Aplicación de QuasiFlow en análisis experimentales

El principal objetivo de QuasiFlow, como ya se ha expuesto, es el estudio de las variables que definen plasticidad genética de las cuasiespecies virales. Para estudiar ésta y otras posibles aplicaciones de QuasiFlow, hemos analizado: (1) coinfecciones de geminivirus y (2) variabilidad de secuencias en DNA mitocondrial (DNAMt) humano, las cuales se muestran a continuación.

2.1. Identificación de moléculas recombinantes en coinfecciones de TYLCV

Para la búsqueda de recombinantes entre geminivirus se tomaron dos muestras de plantas de *S. lycopersicum* coinfectadas simultáneamente con los geminivirus TYLCV TYLCV-Mld, TYLCMaV y TYLCV-IL a 15 y 30 días postinfección (dpi). Para cada una de las muestras se extrajo el DNA total y se amplificó por RCA por duplicado tal como se indica en el apartado 4.1. de Materiales y Métodos. Como paso previo a la secuenciación masiva se realizó un análisis de restricción de las muestras amplificadas por RCA con las endonucleasas *KpnI* y *Clal*. Como se muestra en la **Figura 8A**, *KpnI* tiene una sola diana de corte en el genoma de TYLCMaV, mientras que ese sitio de restricción no existe en

TYLCV-Mld ni en TYLCV-IL. *Clal* (Figura 8B), al contrario, no tiene sitio de corte en el genoma de TYLCMaV, pero presenta una diana en TYLCV-IL y dos en TYLCV-Mld.

Los resultados mostrados en la Figura 8C permiten concluir que la muestra de *S. lycopersicum* a 15 dpi está coinfectada al menos con TYLCMaV y TYLCV-IL, mientras que la muestra de 30 dpi presenta una coinfección de los tres virus.

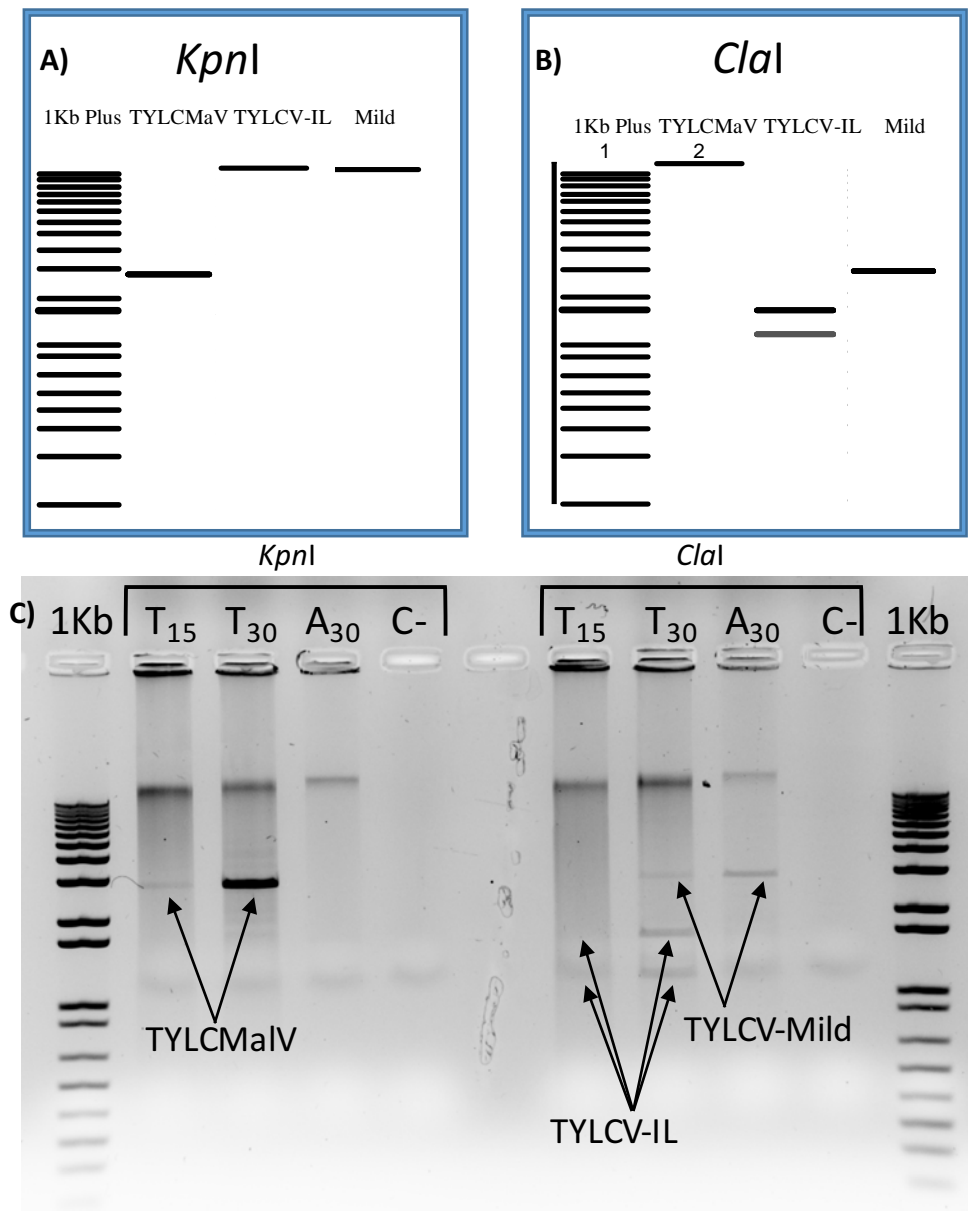


Figura 8. Identificación de las especies virales implicadas en la coinfección por análisis de restricción de concatémeros de TYLCMaV, TYLCV-IL y TYLCV-Mld. Predicción de la digestión con *KpnI* (A) y *Clal* (B). C) Análisis de restricción por electroforesis en gel de agarosa al 1% del producto de RCA de tomate coinfectado a 15dpi (T₁₅), Tomate coinfectado a 30 dpi (T₃₀), *Arabidopsis* mono infectada con TYLCV-Mld a 30dpi (A₃₀) y control de digestión sin DNA (C-).

Los datos de secuenciación Illumina MiSeq se analizaron con QuasiFlow. Para identificar posibles moléculas recombinantes realizamos un primer análisis utilizando la secuencia de TYLCV-Mld como referencia. A partir de las lecturas analizadas, QuasiFlow reconstruyó una misma secuencia consenso para las dos muestras procedentes de *S. lycopersicum* infectadas a 15 y 30 dpi.

La comparación de esta secuencia consenso completa con la secuencia TYLCV-Mld (AF071228) mostró que ambas secuencias compartían un 98% de identidad, indicando la existencia de cambios en la secuencia consenso con respecto a la referencia. Sin embargo, cuando se utilizaron únicamente los nucleótidos 1-976 de la secuencia consenso, se obtuvo una identidad del 100% con TYLCV-IL, mientras que el resto de la secuencia consenso presentaba una identidad del 100% con TYLCMaV. En la Figura 10 se indican los cambios en la secuencia consenso asociados a las secuencias de referencia de TYLCMaV (color rojo) y TYLCV-IL (amarillo). Estos resultados indican que las variante de TYLCV-IL y TYLCMaV son mayoritarias en las coinfecciones respecto a TYLCV-Mld a 15 y 30 dpi, ya que los cambios en la consenso no corresponden a SNPs sino a las diferencias en la secuencia de referencia de ambas especies de TYLCV (**Figura 9**) con la secuencia de referencia de TYLCV-Mld empleada en el análisis.

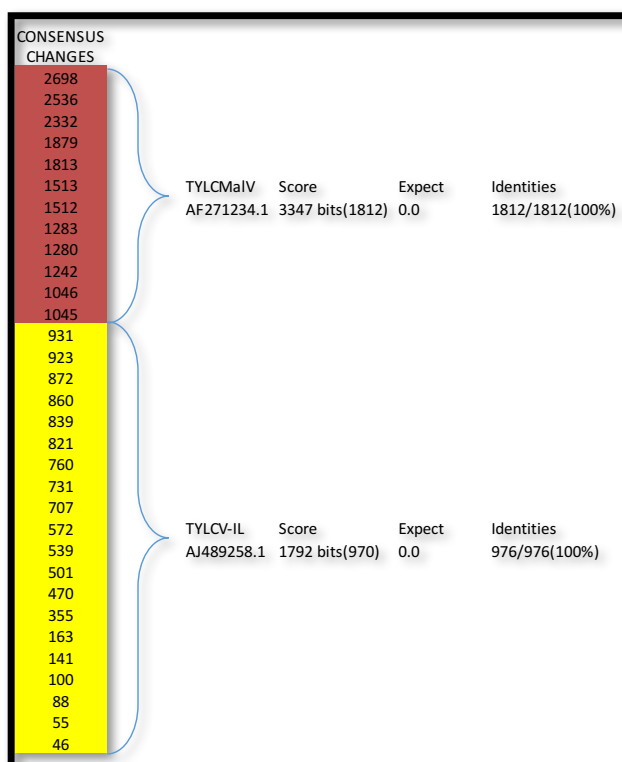


Figura 9. Representación esquemática de los cambios identificados en la secuencia consenso de las muestras de *S. lycopersicum* a 15 y 30 dpi. Se indica la posición del nucleótido en el que se encuentra el cambio con respecto a la secuencia de referencia. En fondo rojo se indican los cambios en la secuencia consenso con respecto a la secuencia de TYLCMaV (AF271234). En fondo amarillo se muestran los cambios en la secuencia consenso con respecto a la secuencia de TYLCV-IL (AJ489258).

Para obtener una representación gráfica de las cuasiespecies virales se diseñó un sistema basado en redes en el que los haplotipos aparecen como nodos y las mutaciones con respecto a la secuencia de referencia aparecen conectadas con flechas a los haplotipos en los que se han detectado. Así, en la **Figura 10**, se muestra la reconstrucción de la cuasiespecie en la muestra de *S. lycopersicum* coinfectada a 15 dpi. Se observa una

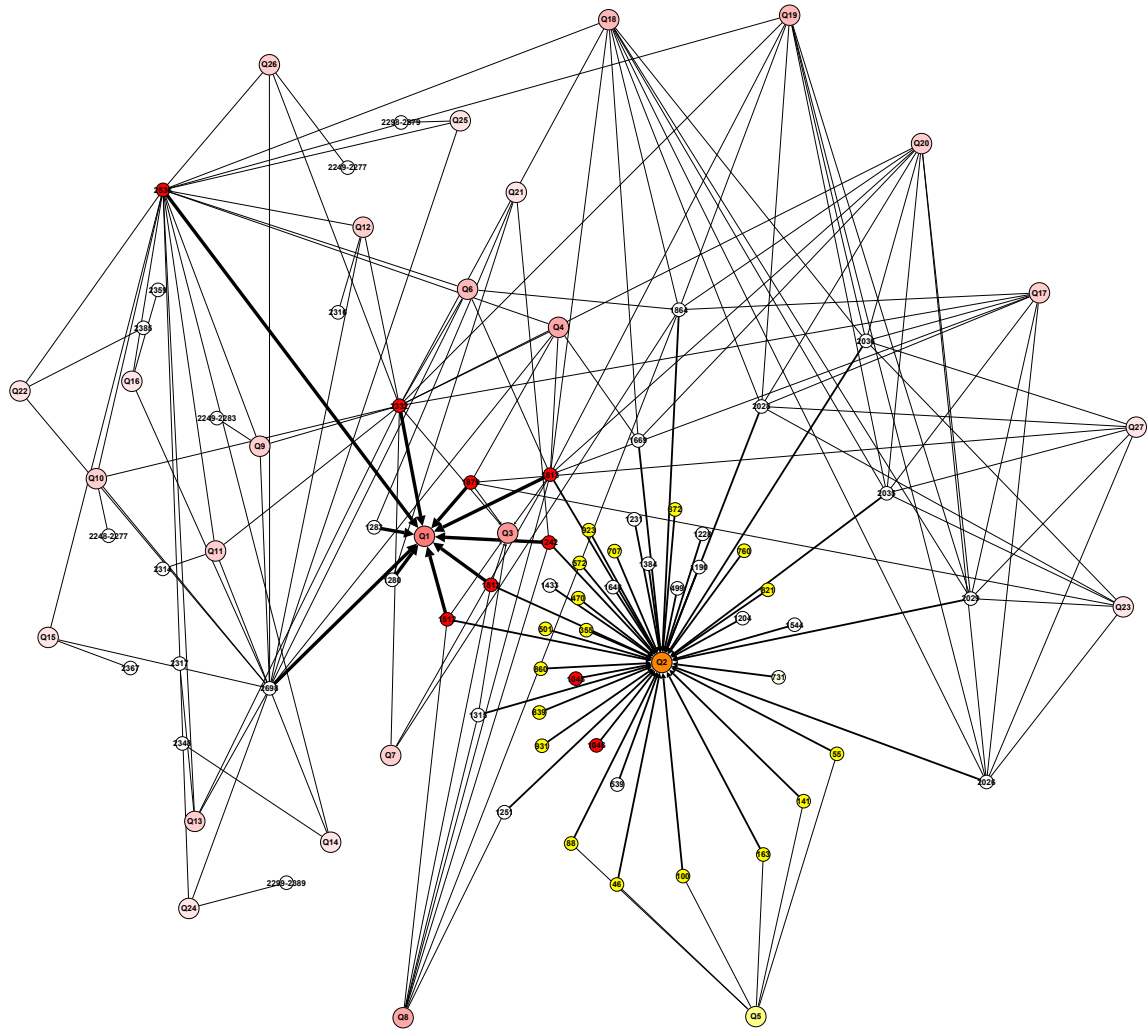


Figura 10. Representación gráfica de la distribución de haplotipos en la cuasiespecie viral de *S. lycopersicum* a 15dpi. Los haplotipos se representan como los círculos de mayor tamaño con la nomenclatura “Q_n” y en un color según los cambios con respecto a la secuencia de TYLCV-MId que se presentan como círculos pequeños con un indicativo numérico que hace referencia a la posición con respecto a la referencia en la que se ha producido el cambio. El grosor de las flechas indica la abundancia relativa de ese cambio dentro de la población viral. El color rojo está asociado a cambios con respecto a la referencia asociados a la secuencia de referencia de TYLCMaV, mientras que el color amarillo está asociado a cambios con respecto a la referencia asociados a la secuencia de referencia de TYLCV-IL. En naranja se representan haplotipos con cambios asociados tanto a TYLCV-IL como a TYLCMaV, que se han interpretado como recombinantes entre ambas especies.

gran mayoría de haplotipos (Q_n) con cambios asociados sólo a TYLCMaV (nodos rojos en los que se indica la posición del cambio dentro de la secuencia referencia de TYLCV-Mld). Tan sólo se detectó un haplotipo con cambios en su secuencia asociado a TYLCV-IL (Q_5). Interesantemente, se observa la existencia de un haplotipo (Q_2) que contiene algunas mutaciones en su secuencia presentes en TYLCMaV y otras en TYLCV-IL, el cual correspondería a un recombinante de ambas especies de TYLCV. Los haplotipos mayoritarios dentro de la cuasiespecie están representados por flechas más gruesas, identificándose de esta forma la secuencia maestra dentro de la población.

En la **Figura 11** se muestra la reconstrucción de la cuasiespecie en la muestra de *S. lycopersicum* coinfectada a 30 dpi.

La cuasiespecie obtenida es más compleja que la analizada a 15 dpi, con un mayor número de haplotipos (H_n color rosa). La mayoría de ellos presentan cambios presentes en TYLCMaV, al igual que lo observado en la muestra a 15 dpi. También se observó un incremento en el número de haplotipos con cambios propios de TYLCV-IL (H_n color amarillo) y en el número de haplotipos recombinantes (H_n color naranja). Al igual que en la cuasiespecie reconstruida a 15 dpi, los haplotipos más representados en la cuasiespecie a 30 dpi fueron los recombinantes, junto con algún haplotipo asociado a TYLCMaV. Estos resultados confirman la coexistencia de TYLCV-Mld, TYLCV-IL y TYLCMaV, tal como se vio en análisis de restricción (figura 8), y además revelan la presencia de recombinantes entre TYLCV-IL y TYLCMaV.

Para estudiar en profundidad la capacidad de Quasiflow para detectar secuencias recombinantes, se repitió el análisis mostrado en las Figuras 10 y 11 tomando como referencia una secuencia artificial de 5563 pb que contiene las secuencias TYLCV-IL y TYLCMaV en tándem (TYLCVIL+TYLCMaV_Híbrid.fasta). El estudio se realizó sobre la muestra a 30 dpi dado que el estudio previo sugiere la existencia de un mayor número de recombinantes a este tiempo de infección.

Los resultados obtenidos se muestran en la **Figura 12** en formato de salida *scatterplot*. Se representan los recombinantes detectados entre moléculas de TYLCV-IL, entre moléculas de TYLCMaV y recombinantes entre ambas moléculas TYLCV-IL y TYLCMaV. Los eventos de recombinación ocurridos entre moléculas de TYLCV-IL se concentran entre la región 3' de cada uno de los genes REP y V2. En cambio, las recombinaciones

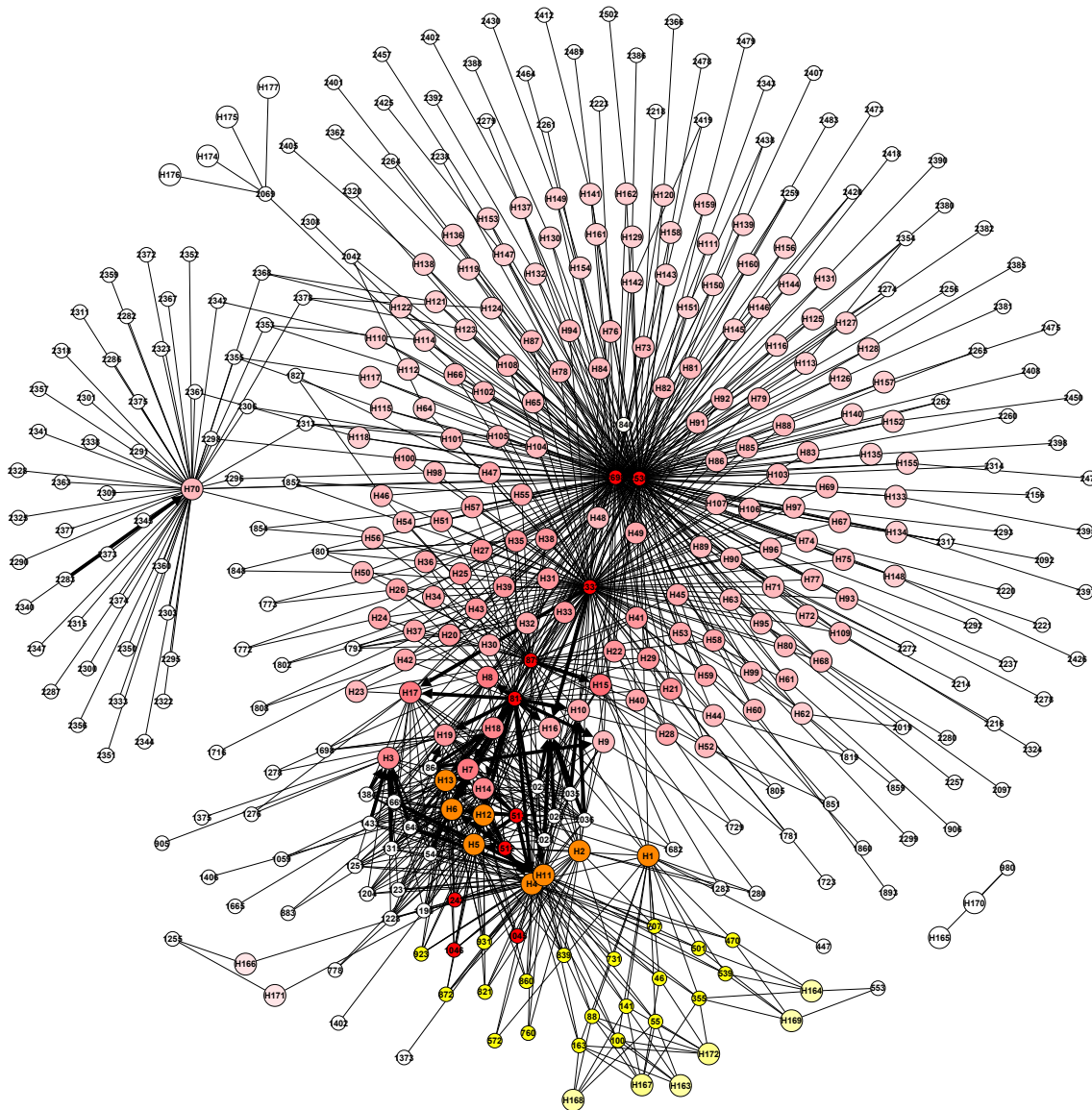


Figura 11. Representación gráfica de la distribución de haplotipos en la cuasiespecie viral de *S. lycopersicum* a 30dpi. Los haplotipos se representan como los círculos de mayor tamaño con la nomenclatura “H_n” y en un color según los cambios con respecto a la secuencia de TYLCV-MId que se presentan como círculos pequeños con un indicativo numérico que hace referencia a la posición con respecto a la referencia en la que se ha producido el cambio. El grosor de las flechas indica la abundancia relativa de ese cambio dentro de la población viral. El color rojo está asociado a cambios con respecto a la referencia asociados a la secuencia de referencia de TYLCMaV, mientras que el color amarillo está asociado a cambios con respecto a la referencia asociados a la secuencia de referencia de TYLCV-IL. En naranja se representan haplotipos con cambios asociados tanto a TYLCV-IL como a TYLCMaV, que se han interpretado como recombinantes entre ambas especies

ocurridas entre moléculas de TYLCMaV no siguen un patrón claro como en TYLCV-IL, observándose una mayor tendencia de la región codificante de REP a recombinar con el

resto del genoma de TYLCV-MaV, y de la región comprendida por los ORFs Ren y TraP a recombinar entre ellos. Las recombinaciones detectadas entre los genomas de TYLCV-IL y TYLCMaV ocurren principalmente en la región de la CP de TYLCV-IL con la correspondiente de TYLCMaV y la región inicial del ORF de REP de TYLCV-IL con dicha región de TYLCMaV. Además también se encontró un cúmulo de eventos de recombinación entre la región 3' de REP y la V2 de TYLCV-IL, tal como ocurría en las recombinaciones de genomas de TYLCV-IL.

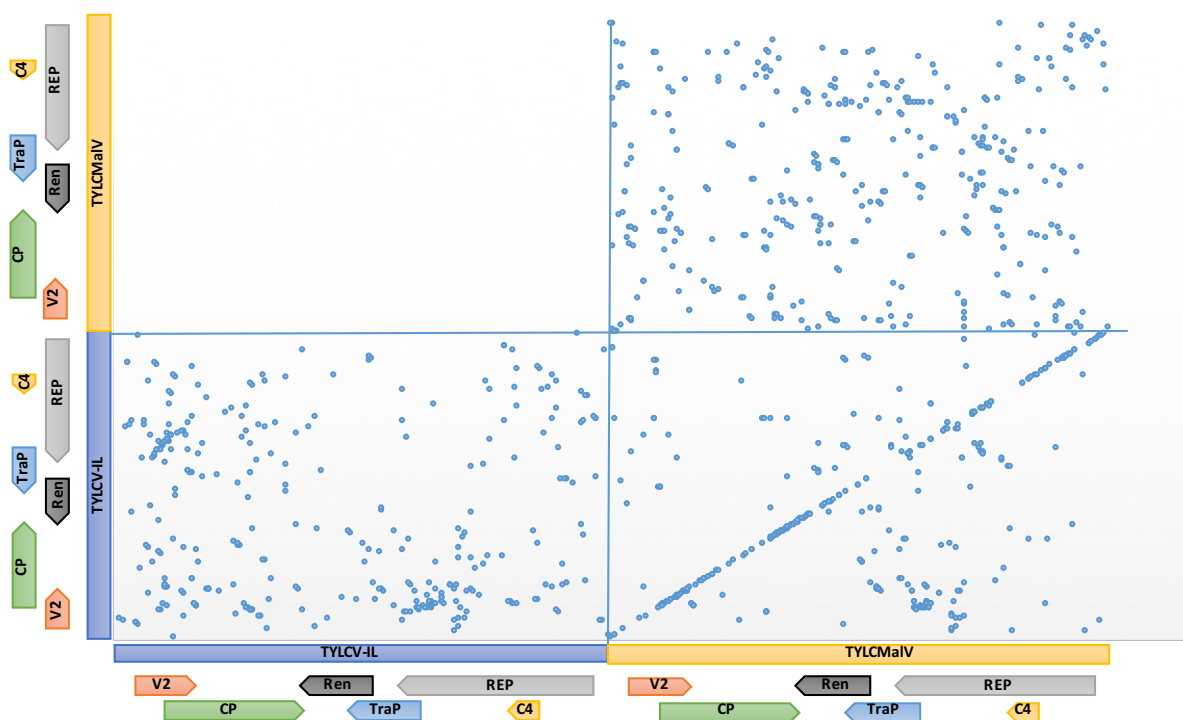


Figura 12. Representación en formato Scatter plot de las recombinaciones halladas en la muestra *S. lycopersicum* a 30dpi. Los ejes X e Y corresponden a las secuencias de TYLCV-IL (azul) y TYLCMaV (amarillo), donde se indica la posición de cada uno de los ORF. Cada cuadrante representa los eventos de recombinación entre ambas especies de geminivirus. El cuadrante superior derecho muestra las recombinaciones encontradas entre moléculas de TYLCMaV. En el cuadrante inferior izquierdo, las recombinaciones asociadas a secuencias de TYLCV-IL y en el cuadrante inferior derecho, las recombinaciones asociadas a la interacción entre ambas especies.

2.2. Caracterización de variantes de nucleótidos con baja frecuencia en DNA mitocondrial humano

Hemos utilizado QuasiFlow para realizar un análisis de variantes de muestras de DNA que presentan una menor frecuencia de mutación que la observada en virus, como es

el DNA mitocondrial. Para ello seleccionamos muestras de DNA mitocondrial de la base de datos de secuenciación masiva SRA del NCBI (obtenidas con Illumina MiSeq, con lecturas pareadas de 150 nt), recogidas en la Tabla 1 de Materiales y métodos, con la intención de verificar la capacidad de análisis del flujo. Las 47 muestras de DNAm seleccionadas fueron tratadas como réplicas biológicas siempre que la procedencia de la muestra y el género de la persona a la que se le extrajo fuesen los mismos: 7 y 8 muestras pertenecían a DNA de línea celular linfoblastoide de hombre y de mujer respectivamente, 5 muestras obtenidas de saliva de hombre, y 3 y 24 muestras de DNA extraído de sangre de hombre y de mujer respectivamente (Tabla 1 de Materiales y métodos).

La secuencia del genoma de DNAm humano utilizada como referencia en el análisis fue la denominada *Revised Cambridge Reference Sequence* ("rCRS"), de un tamaño de 16569 nts con ID de GenBank NC_012920. Esta secuencia es la usada como secuencia estándar en el análisis comparativo de DNAm humano.

El tipo de salida que genera QuasiFlow de forma automática corresponde a las **Tablas 3-7 y Figuras 13-19**.

- **Caracterización de la variabilidad genética en muestras de DNA mitocondrial: selección calidad lecturas en estudio**

Las lecturas de secuenciación masiva SRA de las 47 muestras indicadas anteriormente fueron analizadas con QuasiFlow. Tras aplicar los filtros correspondientes, el número de lecturas de cada muestra fue bastante homogéneo, oscilando entre las $1,38 \times 10^5$ lecturas para la réplica 4 de DNA de línea celular linfoblastoide de mujer y las $4,32 \times 10^5$ lecturas correspondientes a la réplica 10 de la misma muestra. Debido a que el *coverage* medio por nucleótido de las muestras no superó los 1500X, no se tuvieron en cuenta las variantes con frecuencias de aparición menores a 1×10^{-3} cambios por posición nucleotídica (0,001%). En la Tabla 3 se muestra un resumen de los resultados crudos extraídos del análisis automatizado del DNA mitocondrial con QuasiFlow.

Tabla 3. Tabla resumen del análisis de las muestras mitocondriales en estudio

Nombre	Genero	Extracción	Lecturas	Nt secuenciados
r1_Female_Blood	Hembra	Sangre	3,47E+05	5,16E+07
r10_Female_Blood	Hembra	Sangre	1,74E+05	2,58E+07
r11_Female_Blood	Hembra	Sangre	2,05E+05	3,04E+07
r12_Female_Blood	Hembra	Sangre	2,04E+05	3,03E+07
r13_Female_Blood	Hembra	Sangre	2,44E+05	3,62E+07
r14_Female_Blood	Hembra	Sangre	2,13E+05	3,16E+07
r15_Female_Blood	Hembra	Sangre	3,30E+05	4,89E+07
r16_Female_Blood	Hembra	Sangre	2,65E+05	3,92E+07
r17_Female_Blood	Hembra	Sangre	3,49E+05	5,17E+07
r18_Female_Blood	Hembra	Sangre	3,37E+05	5,00E+07
r19_Female_Blood	Hembra	Sangre	2,89E+05	4,29E+07
r2_Female_Blood	Hembra	Sangre	1,96E+05	2,90E+07
r20_Female_Blood	Hembra	Sangre	3,48E+05	5,16E+07
r21_Female_Blood	Hembra	Sangre	3,54E+05	5,26E+07
r22_Female_Blood	Hembra	Sangre	1,95E+05	2,88E+07
r23_Female_Blood	Hembra	Sangre	2,14E+05	3,18E+07
r24_Female_Blood	Hembra	Sangre	1,58E+05	2,34E+07
r3_Female_Blood	Hembra	Sangre	2,47E+05	3,66E+07
r4_Female_Blood	Hembra	Sangre	2,15E+05	3,19E+07
r5_Female_Blood	Hembra	Sangre	2,22E+05	3,29E+07
r6_Female_Blood	Hembra	Sangre	3,49E+05	5,18E+07
r7_Female_Blood	Hembra	Sangre	2,92E+05	4,34E+07
r8_Female_Blood	Hembra	Sangre	2,73E+05	4,05E+07
r9_Female_Blood	Hembra	Sangre	1,98E+05	2,93E+07
r1_Female_Lymphoblastoid	Hembra	Linfoblastos	3,47E+05	5,15E+07
r10_Female_Lymphoblastoid	Hembra	Linfoblastos	4,32E+05	6,41E+07
r2_Female_Lymphoblastoid	Hembra	Linfoblastos	3,35E+05	4,97E+07
r3_Female_Lymphoblastoid	Hembra	Linfoblastos	3,56E+05	5,29E+07
r4_Female_Lymphoblastoid	Hembra	Linfoblastos	1,38E+05	2,05E+07
r6_Female_Lymphoblastoid	Hembra	Linfoblastos	2,38E+05	3,53E+07
r7_Female_Lymphoblastoid	Hembra	Linfoblastos	3,04E+05	4,51E+07
r8_Female_Lymphoblastoid	Hembra	Linfoblastos	3,50E+05	5,19E+07
r2_Male_Blood	Varón	Sangre	2,94E+05	4,37E+07
r3_Male_Blood	Varón	Sangre	3,25E+05	4,82E+07
r4_Male_Blood	Varón	Sangre	1,89E+05	2,80E+07
r1_Male_Lymphoblastoid	Varón	Linfoblastos	2,43E+05	3,61E+07
r2_Male_Lymphoblastoid	Varón	Linfoblastos	2,28E+05	3,38E+07
r3_Male_Lymphoblastoid	Varón	Linfoblastos	2,70E+05	4,00E+07
r4_Male_Lymphoblastoid	Varón	Linfoblastos	2,07E+05	3,08E+07
r5_Male_Lymphoblastoid	Varón	Linfoblastos	3,06E+05	4,54E+07

r6_Male_Lymphoblastoid	Varón	Linfoblastos	3,20E+05	4,74E+07
r7_Male_Lymphoblastoid	Varón	Linfoblastos	3,17E+05	4,71E+07
r1_Male_Saliva	Varón	Saliva	2,28E+05	3,39E+07
r2_Male_Saliva	Varón	Saliva	2,67E+05	3,96E+07
r3_Male_Saliva	Varón	Saliva	3,49E+05	5,18E+07
r4_Male_Saliva	Varón	Saliva	3,51E+05	5,21E+07
r5_Male_Saliva	Varón	Saliva	3,16E+05	4,69E+07

- **Caracterización de la variabilidad genética en muestras de ADN mitocondrial: tipo y frecuencia de mutaciones**

Para caracterizar la variabilidad genética de las muestras de ADN mitocondrial (DNAMt) citadas se analizaron los SNPs, la frecuencia de SNPs e InDels y los puntos de recombinación.

En primer lugar, se realizó una comparación de la variabilidad genética (SNPs) entre las muestras en función de su procedencia (origen celular y género). Los resultados de este estudio (**Figura 13**) confirman la existencia de cambios con respecto a la secuencia de referencia en todas las muestras analizadas, que oscilan entre un mínimo de 12 y un

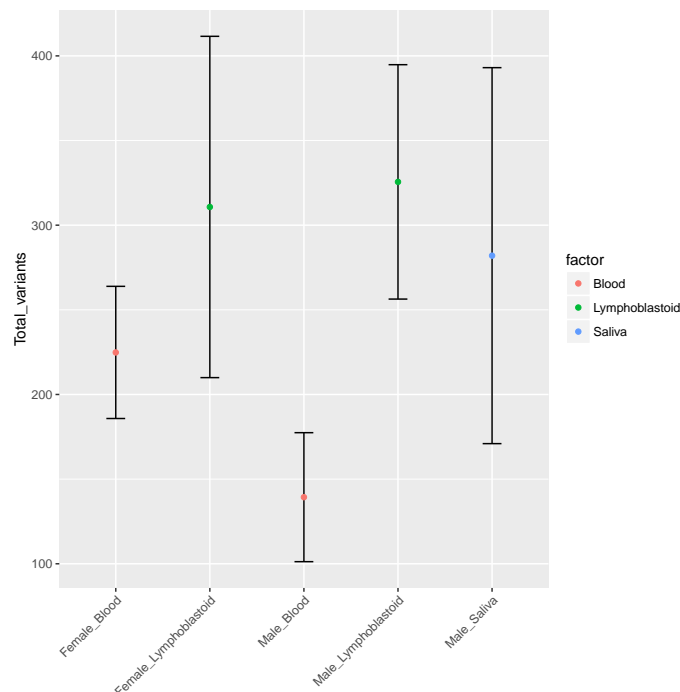


Figura 13. Cálculo del número de variantes en secuencias de DNA mitocondrial con respecto al tipo muestra (origen celular y género). En rojo se ha representado las muestras de sangre, en verde las muestras de líneas celulares linfoblastoides y en azul las muestras procedentes de saliva. Las barras de errores corresponden a la desviación estándar entre réplicas biológicas. No se observaron diferencias significativas tras el análisis ANOVA realizado por QuasiFlow.

máximo de 863 cambios por genoma. Sin embargo, el análisis ANOVA realizado por QuasiFlow (P-valor <0,05) no detecta diferencias significativas entre las muestras.

En cuanto al tipo de cambios encontrados (**Tabla 4**), se aprecia un sesgo hacia las transiciones, siendo A→G y T→C las más frecuentes en todas las muestras analizadas, seguidas de C→T y A→G. Por su parte, las transversiones fueron muy escasas, de hecho en 12 de las 47 muestras no apareció ninguna. En la **Tabla 4** se encuentra la comparación entre muestra origen de ADNmt y la ratio Ts/Tv en aquellas muestras en las que se detectaron transversiones. Los resultados indican que las secuencias de DNA mitocondrial procedentes de linfoblastos y de saliva de hombre presentan un menor ratio Ts/Tv, contrariamente a lo que ocurrió con el número de variantes, siendo ambos valores menores a 50. Las muestras restantes procedentes de sangre de mujer y de hombre y de linfoblastos de mujer presentaron ratios Ts/Tv cercanos a 100 (**Tabla 4**).

La frecuencia en la que se encuentran las variantes dentro de las moléculas de ADNmt determina la existencia de heteroplasmas, en el que variantes diferentes del ADNmt que coexisten en un mismo organismo, y homoplasmas, en el caso de que todas las copias de ADNmt sean idénticas entre sí (ya sean variantes nucleotídicas con respecto a la secuencia de referencia rCRS o no). Su estudio es importante dado que determinadas heteroplasmas y homoplasmas están asociadas en humanos a la aparición de enfermedades mitocondriales o cáncer. QuasiFlow generó una Tabla con todas las variantes encontradas en cada una de las 47 muestras y su porcentaje dentro del total de la muestra (resultados no mostrados en la presente Tesis) A modo de ejemplo, en este apartado sólo se indica el estudio correspondiente a la muestra 1 de ADNmt procedente de la línea celular linfoblastoide de mujer (**Tabla 5**), donde se indican las 26 homoplasmas detectadas para esta muestra junto con sus porcentajes de ocurrencia dentro del total de lecturas analizadas. Una consulta manual en la web mitomap.com sobre enfermedades asociadas a estos cambios, mostró que 8 homoplasmas detectadas están asociadas a cáncer de ovario y/o tiroides. Estas variantes fueron las encontradas en las posiciones 263 (A→G) (región hipervariable 2), 16519 (T→C) (región D-loop), 6473 (C→T) (COI), 499 (G→A) (región hipervariable 3), 827(A→G) (RNAr 12S), 16217 (T→C) (*d-loop*), 73 (A→G) (región hipervariable 2) y 16189 (T→C) (*d-loop*). El resto de homoplasmas encontradas en esta muestra también estaban documentadas en la base

de datos Mitomap pero ninguna de ellas está asociada a enfermedad. El número de heteroplasmas que fueron encontradas en esta muestra fue de 310, de las cuales sólo 10 superaban una frecuencia del 1%, mientras que el resto mostraron un porcentaje inferior al 1%.

Nombre	SNPs	SNPs_rel	G→A	A→G	T→C	C→T	C→A	T→G	G→C	A→C	T→A	G→T	A→T	C→G	Transiciones	Transversiones	Ts/Tv ^a	InDels
r1_Female_Blood	338	6,55E-06	44	115	128	45	2	0	1	1	1	0	0	0	332	5	66,4	17
r10_Female_Blood	838	3,24E-05	103	122	275	245	26	1	1	24	15	1	20	4	745	92	8,1	16
r11_Female_Blood	116	3,81E-06	12	36	52	12	2	0	0	0	0	0	1	0	112	3	37,33	5
r12_Female_Blood	87	2,88E-06	8	30	41	7	0	0	0	0	0	0	0	0	86	0	N.A. ^b	11
r13_Female_Blood	104	2,87E-06	12	37	47	7	0	0	0	0	0	0	0	0	103	0	N.A.	10
r14_Female_Blood	72	2,28E-06	9	20	32	10	0	0	0	0	0	0	0	0	71	0	N.A.	11
r15_Female_Blood	311	6,36E-06	17	119	157	15	0	0	0	0	1	0	1	0	308	2	154	14
r16_Female_Blood	141	3,59E-06	11	48	68	13	0	0	0	0	0	0	0	0	140	0	N.A.	12
r17_Female_Blood	261	5,05E-06	19	78	138	24	1	0	0	0	0	0	0	0	259	1	259	15
r18_Female_Blood	158	3,16E-06	15	55	67	19	0	0	1	0	0	0	0	0	156	1	156	13
r19_Female_Blood	236	5,50E-06	16	87	113	19	0	0	0	0	0	0	0	0	235	0	N.A.	15
r2_Female_Blood	101	3,48E-06	7	32	48	11	1	0	0	0	0	0	1	0	98	2	49	9
r20_Female_Blood	201	3,89E-06	20	68	94	18	0	0	0	0	0	0	0	0	200	0	N.A.	16
r21_Female_Blood	331	6,29E-06	24	107	178	17	2	0	0	0	1	0	1	0	326	4	81,5	19
r22_Female_Blood	99	3,43E-06	7	34	45	11	0	0	0	0	0	0	1	0	97	1	97	8
r23_Female_Blood	69	2,17E-06	10	18	26	12	1	0	0	0	0	0	1	0	66	2	33	5
r24_Female_Blood	68	2,90E-06	5	27	29	5	0	1	0	0	0	0	0	0	66	1	66	7
r3_Female_Blood	175	4,78E-06	8	60	95	11	0	0	0	0	0	0	0	0	174	0	N.A.	11
r4_Female_Blood	214	6,71E-06	7	85	114	7	0	0	0	0	0	0	0	0	213	0	N.A.	9
r5_Female_Blood	96	2,92E-06	8	41	39	6	0	0	0	0	0	0	0	1	94	1	94	11
r6_Female_Blood	707	1,37E-05	24	301	352	28	1	0	0	0	0	0	0	0	705	1	705	17

Nombre	SNPs	SNPs_rel	G→A	A→G	T→C	C→T	C→A	T→G	G→C	A→C	T→A	G→T	A→T	C→G	Transiciones	Transversiones	Ts/Tv ^a	InDels
r7_Female_Blood	329	7,58E-06	15	130	158	24	1	0	0	0	0	0	0	0	327	1	327	8
r8_Female_Blood	236	5,83E-06	12	91	117	15	0	0	0	0	0	0	0	0	235	0	N.A.	15
r9_Female_Blood	132	4,50E-06	13	40	65	11	0	0	0	0	0	1	1	0	129	2	64,5	10
r1_Female_Lymphoblastoid	279	5,42E-06	21	94	142	18	1	2	0	0	0	0	0	0	275	3	91,67	17
r10_Female_Lymphoblastoid	623	9,72E-06	84	117	216	158	11	1	2	11	7	1	10	4	575	47	12,23	19
r2_Female_Lymphoblastoid	250	5,03E-06	23	88	121	16	1	0	0	0	0	0	0	0	248	1	248	15
r3_Female_Lymphoblastoid	140	2,65E-06	15	50	60	13	0	0	1	0	0	0	0	0	138	1	138	11
r4_Female_Lymphoblastoid	12	5,86E-07	0	6	2	3	0	0	0	0	0	0	0	0	11	0	N.A.	3
r6_Female_Lymphoblastoid	170	4,82E-06	21	57	69	21	1	0	0	0	0	0	0	0	168	1	168	10
r7_Female_Lymphoblastoid	863	1,91E-05	114	136	266	251	27	1	2	24	15	2	19	5	767	95	8,07	22
r8_Female_Lymphoblastoid	157	3,02E-06	19	44	75	15	1	0	0	0	2	0	0	0	153	3	51	15
r2_Male_Blood	163	3,73E-06	15	55	80	10	0	0	0	1	0	0	1	0	160	2	80	17
r3_Male_Blood	192	3,98E-06	23	63	92	12	0	0	0	0	0	1	0	0	190	1	190	16
r4_Male_Blood	66	2,36E-06	6	21	28	7	1	1	0	0	0	0	1	0	62	3	20,67	7
r1_Male_Lymphoblastoid	253	7,02E-06	27	45	105	62	5	0	1	1	3	1	0	2	239	13	18,38	11
r2_Male_Lymphoblastoid	87	2,57E-06	10	27	30	17	0	0	1	1	0	0	0	0	84	2	42	7
r3_Male_Lymphoblastoid	612	1,53E-05	72	105	196	176	17	1	2	15	10	0	14	3	549	62	8,85	15
r4_Male_Lymphoblastoid	165	5,36E-06	17	47	69	29	0	0	0	1	1	0	0	0	162	2	81	9
r5_Male_Lymphoblastoid	279	6,15E-06	35	68	117	51	1	0	0	0	4	1	0	1	271	7	38,71	15
r6_Male_Lymphoblastoid	428	9,02E-06	50	89	181	87	5	0	0	4	5	1	3	2	407	20	20,35	15

Nombre	SNPs	SNPs_rel	G→A	A→G	T→C	C→T	C→A	T→G	G→C	A→C	T→A	G→T	A→T	C→G	Transiciones	Transversiones	Ts/Tv ^a	InDels
r7_Male_Lymphoblastoid	462	9,81E-06	64	75	166	115	12	0	1	11	7	0	8	2	420	41	10,24	19
r1_Male_Saliva	69	2,04E-06	8	22	34	4	0	0	0	0	0	0	0	0	68	0	N.A.	11
r2_Male_Saliva	124	3,13E-06	16	43	48	15	0	0	0	1	0	0	0	0	122	1	122	14
r3_Male_Saliva	293	5,66E-06	22	104	141	25	0	0	0	0	0	0	0	0	292	0	N.A.	18
r4_Male_Saliva	409	7,85E-06	41	77	128	154	2	0	0	1	1	1	2	1	400	8	50	15
r5_Male_Saliva	530	1,13E-05	73	101	185	126	14	0	2	10	6	1	8	3	485	44	11,02	28

^b Ts/Tv: ratio transiciones/transversiones

^aN.A.: Muestra no analizada

- **Caracterización de la variabilidad genética en muestras de ADN: análisis iterativo de variantes**

Hemos comparado con QuasiFlow el número del total de variantes del ADNmt de las 47 muestras con respecto al total de lecturas de cada muestra. Para ello, complementamos QuasiFlow con el *script iterative_varscan.rb* (descrito en el apartado Materiales y métodos). Los resultados muestran (**Figura 14**) que el número de variantes encontradas es independiente del número de lecturas ($P > 0,001$ y $R^2 = 0,178$) en las muestras de partida.

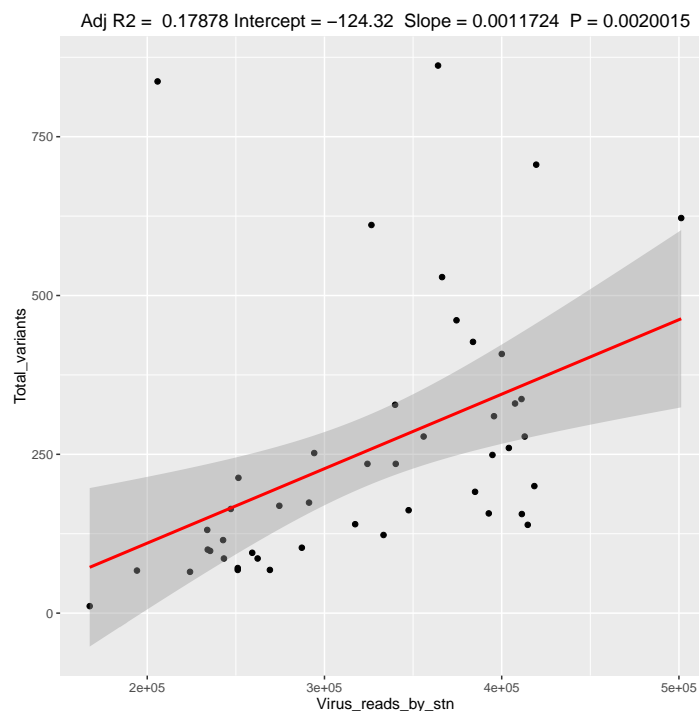


Figura 14. Estudio de la relación entre el número de variables y número de lecturas. Representación gráfica obtenida por Quasiflow. Cada punto representa una muestra de DNAm a partir de los valores obtenidos por QuasiFlow en el bloque 1.

Este resultado nos permitió abordar el análisis iterativo de variantes en las 47 muestras de partida. En la **Figura 15** se muestra el análisis comparativo de la frecuencia de aparición de cambios en cada posición del genoma mitocondrial con respecto a la procedencia de la muestra en estudio. Los resultados muestran una acumulación de variantes de baja frecuencia en todas las muestras, independientemente de la procedencia del material genético (Figura 15), distribuyéndose las heteroplasmas en

valores de frecuencia comprendidos entre el 0,002% y el 0,02%. A partir de este valor de 0,02%, el número de variantes disminuye, no encontrándose ninguna muestra cuyas variantes con una frecuencia superior al 0,02% represente el 0,5% del total de variantes de dicha muestra. También se observó una alta variabilidad, con valores entre 0,002% y 0,02%, en el DNAmT procedente de muestras que comparten un mismo origen celular. Este dato sugiere que la acumulación de mutaciones no guarda relación con el origen celular de la muestra.

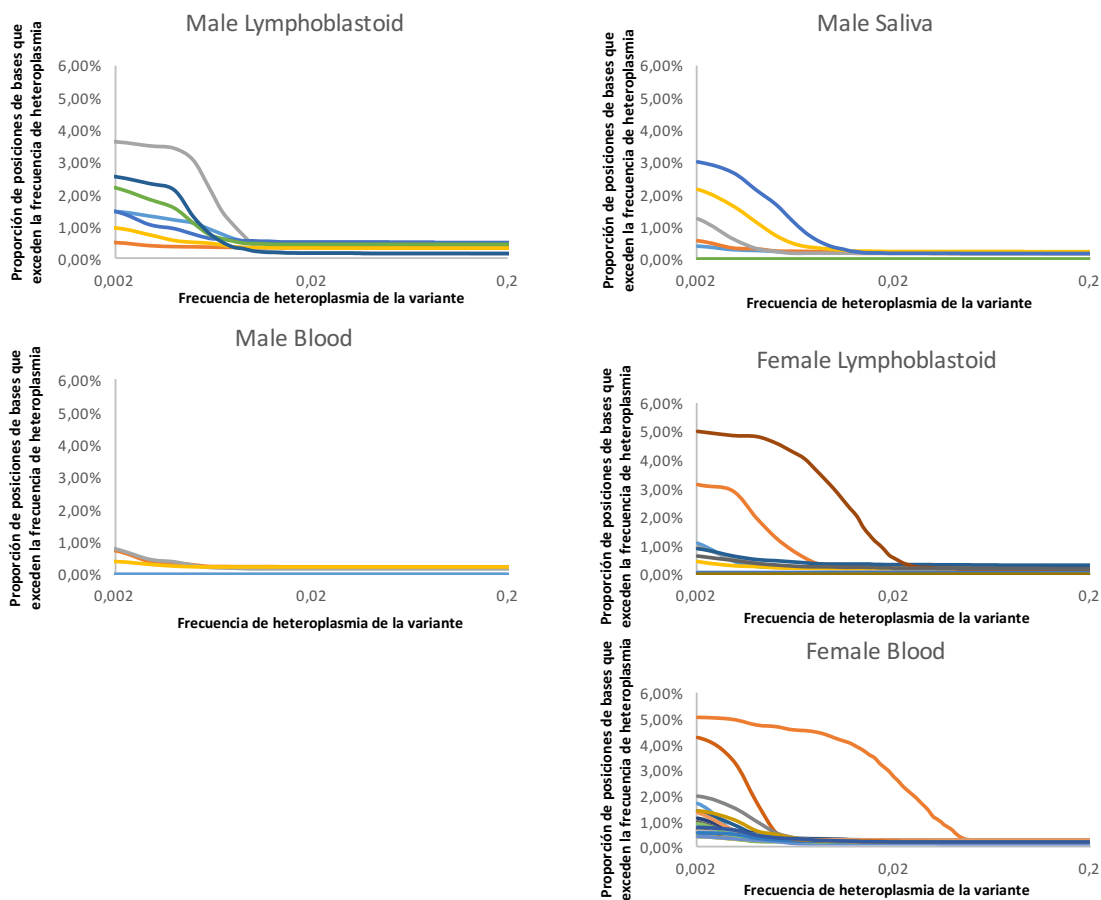


Figura 15. Comparación de variantes detectadas en las muestras analizadas de DNA mitocondrial. Las Gráficas representan la frecuencia de heteroplásmias en relación a la proporción de posiciones (número de posiciones mutadas sobre el total de posiciones nucleotídicas de la referencia) de base que exceden dicha frecuencia, de forma que se pueda apreciar la distribución de variantes con respecto a su frecuencia acumulada.

- **Caracterización de la variabilidad genética en muestras de DNAm: análisis de InDels**

Hemos analizado los InDels de cada una de las 47 muestras de DNAm. A modo de ejemplo, detallamos los InDels identificados en la muestra correspondiente a la muestra 1 de DNAm procedente de línea celular linfoblastoide de mujer (**Tabla 6**). De los 17 InDels hallados, tan sólo 1 corresponde a una inserción, que tiene lugar en la posición 65 (T→TG) y ocurre con una frecuencia del 77%; los 16 InDels restantes correspondieron a deleciones, de las cuales tan sólo 1 mostró una frecuencia de aparición superior al 90%. Esta deleción en el nt 3106 (C→Del) es un artefacto introducido al usar la secuencia de referencia que contiene una N en la posición 3107 para poder conservar la numeración de la versión anterior del genoma de referencia. El resto de deleciones presentaron una frecuencia inferior al 2%, llegando a detectar deleciones con una frecuencia del 0,15%. Estos porcentajes inferiores a los hallados en los SNPs podrían indicar el carácter deletéreo que pueden tener los InDels en el DNAm.

Tabla 6. Tabla de InDels identificados en la réplica 1 de la línea celular linfoblastoide de mujer

Posición	Sec. Referencia	InDel	Frecuencia
65	TG	T	0,38%
302	A	AC	0,60%
513	GCA	G	1,37%
2456	TA	T	0,46%
3105	AC	A	91,77%
4604	CA	C	1,07%
5746	GA	G	0,66%
6220	CT	C	0,30%
6691	GA	G	0,75%
7709	CT	C	0,17%
8094	TA	T	0,12%
8151	CG	C	0,17%
12417	CA	C	3,53%
16183	A	AC	13,37%
16187	CCT	C	0,59%
16188	CT	C	12,29%
16374	AC	A	0,36%

- **Caracterización de la variabilidad genética en muestras de ADNmt: análisis de secuencias recombinantes**

Debido al daño oxidativo al que están sujetas las moléculas de ADNmt dentro de las mitocondrias, éstas podrían ser susceptibles de sufrir procesos de rotura dando lugar a la formación de genomas recombinantes durante el proceso de reparación (Ladoukakis et al., 2011). Sin embargo, hay muy pocos casos descritos de ocurrencia de eventos de recombinación mitocondrial en humanos, aunque en levaduras, hongos y moluscos sí han sido descritos recientemente (Fritsch et al., 2014; Xu et al., 2013; Zbawicka et al., 2014). Por ello hemos estudiado la posible existencia de secuencias recombinantes en las 47 muestras de DNAm_t utilizando la herramienta QuasiFlow. Se muestran los resultados obtenidos para la réplica 1 de la línea celular linfoblastoide de mujer (**Tabla 7**). Los resultados obtenidos muestran la existencia de secuencias recombinantes dentro del DNAm_t en todas las muestras analizadas, siendo mayor la proporción de secuencias recombinantes que implicaron a la cadena de DNA de igual polaridad con respecto a las recombinaciones que implicaron a ambas cadenas de DNA. En cuanto a la distribución de puntos de recombinación, se observó que existen sitios preferentes de recombinación tanto de una sola cadena como de cadena doble.

Tabla 7. Tabla de recombinantes identificados en la réplica 1 de la línea celular linfoblastoide de mujer

Desde	hasta	Repeticiones	Cadenas	Desde	hasta	Repeticiones	Cadenas
1399	1406	1	Doble	16238	15585	1	Simple
6766	6802	1	Doble	14896	15516	1	Simple
8497	5674	1	Doble	11184	15552	1	Simple
16035	16070	1	Doble	8123	1717	1	Simple
1200	1240	1	Doble	435	10916	1	Simple
7764	7784	1	Doble	566	16256	1	Simple
13327	13348	1	Doble	5204	15633	1	Simple
7635	7655	1	Doble	963	15699	1	Simple
5674	8497	1	Doble	7420	15425	1	Simple
623	726	1	Doble	16073	8636	1	Simple
13447	14750	1	Doble	7246	15543	1	Simple
5982	5994	1	Doble	757	8673	1	Simple
2030	16038	1	Doble	7797	15641	1	Simple
2690	2698	1	Doble	469	9221	1	Simple
4955	4995	1	Doble	5429	9238	1	Simple
16106	4386	1	Doble	5444	7862	1	Simple
16266	15542	2	Simple	7471	15682	1	Simple
3755	15629	2	Simple	597	15603	1	Simple
4695	15565	2	Simple	7956	15657	1	Simple
7781	15630	1	Simple	12482	7988	1	Simple
369	15597	1	Simple	14490	3483	1	Simple
7742	15702	1	Simple	5501	15651	1	Simple
467	8940	1	Simple	612	16348	1	Simple
16190	9060	1	Simple	15336	15566	1	Simple
12708	9041	1	Simple	5460	15552	1	Simple
595	9033	1	Simple	5662	15708	1	Simple
14366	9350	1	Simple	14021	11041	1	Simple
8124	15580	1	Simple	13386	3236	1	Simple
7944	15575	1	Simple	5329	11115	1	Simple
1154	3713	1	Simple	1509	16178	1	Simple
6214	9887	1	Simple	5447	8558	1	Simple
4817	9131	1	Simple	14389	8406	1	Simple
8079	15540	1	Simple	5460	8291	1	Simple
507	11864	1	Simple	5722	191	1	Simple
14488	703	1	Simple	14001	8723	1	Simple
7109	15576	1	Simple	5454	7807	1	Simple
13988	1039	1	Simple	13947	14280	1	Simple
8117	15594	1	Simple	16423	15655	1	Simple
5485	15585	1	Simple	597	11996	1	Simple
6419	5660	1	Simple	8106	15581	1	Simple

Desde	hasta	Repeticiones	Cadenas	Desde	hasta	Repeticiones	Cadenas
7722	15821	1	Simple	10857	5360	1	Simple
1155	9694	1	Simple	8983	1653	1	Simple
15424	7419	3	Simple	15646	5313	1	Simple
15448	6492	2	Simple	15546	3578	1	Simple
16181	489	2	Simple	15386	7869	1	Simple
15415	7410	1	Simple	5409	2835	1	Simple
15570	14103	1	Simple	16128	227	1	Simple
15539	7407	1	Simple	8599	7087	1	Simple
15458	7689	1	Simple	15421	7416	1	Simple
7656	7858	1	Simple	7740	5191	1	Simple
8464	7131	1	Simple	8564	1089	1	Simple
7711	5198	1	Simple	3792	963	1	Simple
15460	12326	1	Simple	7653	5006	1	Simple
15469	5417	1	Simple	13604	1019	1	Simple
16083	869	1	Simple	15426	14653	1	Simple
14704	2384	1	Simple	15546	5389	1	Simple
15404	15977	1	Simple	15429	7424	1	Simple
15476	11528	1	Simple	16142	492	1	Simple
15405	14355	1	Simple	8965	5146	1	Simple
7697	5224	1	Simple	15391	5446	1	Simple
1746	1257	1	Simple	15605	377	1	Simple
8289	8281	1	Simple	15403	5328	1	Simple
2693	13268	1	Simple	494	2416	1	Simple
15481	7967	1	Simple	1040	13989	1	Simple
7693	321	1	Simple	15415	3375	1	Simple
16194	967	1	Simple	15465	5201	1	Simple
7721	7413	1	Simple	15616	7080	1	Simple
9151	5336	1	Simple	7329	7345	2	Doble
15462	5494	1	Simple	6671	7356	1	Doble
8766	438	1	Simple	7751	7781	1	Doble
15404	7400	1	Simple	6750	14482	1	Doble
15425	7677	1	Simple	13974	649	1	Doble
8579	13331	1	Simple	5991	6031	1	Doble
15638	5368	1	Simple	637	628	1	Doble
8434	15647	1	Simple	2925	6870	1	Doble
9205	450	1	Simple	7607	1691	1	Doble
8516	373	1	Simple	3689	3667	1	Doble
16172	5292	1	Simple	6980	7714	1	Doble
1730	7539	1	Simple	1379	1374	1	Doble
16173	482	1	Simple	13929	7910	1	Doble
15925	816	1	Simple	7602	7547	1	Doble
7683	5357	1	Simple	7969	7944	1	Doble

Desde	hasta	Repeticiones	Cadenas	Desde	hasta	Repeticiones	Cadenas
5443	7510	1	Doble	3520	7614	1	Doble
6631	6419	1	Doble	7799	6934	1	Doble
15387	15379	1	Doble	7936	8028	1	Doble
7269	7910	1	Doble	431	6921	1	Doble
8019	8008	1	Doble	4262	4242	1	Doble
6021	6001	1	Doble	2672	2607	1	Doble
9777	9741	1	Doble	7510	5443	1	Doble

A modo de ejemplo describimos el análisis de las secuencias recombinantes halladas en el DNAmT de las muestras de sangre de mujer, de la cuales disponemos de un mayor número de muestras. Los resultados muestran un patrón muy diferente entre las recombinaciones que implican a una cadena o a las dos cadenas (**Figura 16**). En cuanto a las recombinaciones en una cadena, se observaron tres regiones de recombinación claramente delimitadas: la región del DNAmT que comprende los nt 7500 a 8000 tiene tendencia a recombinar con cualquier región del genoma. La región entre los nt 1 a 1000

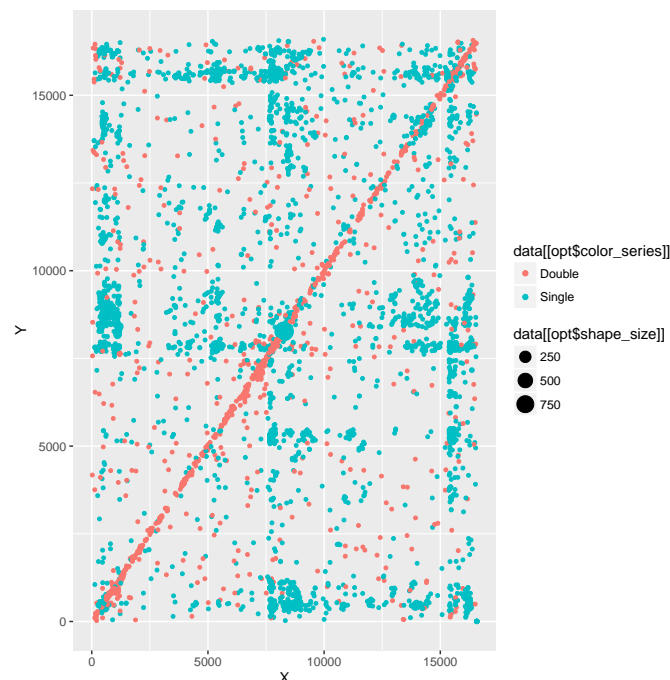


Figura 16. Representación en formato Scatter plot de recombinaciones encontradas en las muestras extraídas de DNA mitocondrial de sangre de mujer. Los valores de ambos ejes son posiciones nucleotídicas de la secuencia de DNAmT de referencia rCRS. El grosor de los puntos indica el número de veces que se repite esa recombinación y el color rojo y azul indican si la recombinación ocurre entre regiones de la misma cadena o entre cadenas complementarias respectivamente.

recombina de forma clara con la región entre los nt 7500 a 16000 y por último, la región entre los nt 15000 y 16000 recombina con la región 1 a 7500. En el caso de las recombinaciones que implican a dos cadenas, se observó un patrón aleatorio de recombinaciones a lo largo de todo el genoma pero con tendencia a producir recombinaciones entre zonas del genoma muy cercanas entre sí. Este patrón no se observó en las recombinaciones que implican a una cadena, cuya tendencia como se ha expuesto aquí, fue producir recombinaciones entre zonas alejadas del genoma.

- **Caracterización de la variabilidad genética en muestras de DNA mitocondrial: análisis multivariante**

Hasta el momento, todos los análisis realizados sobre las muestras de DNA mitocondrial sugieren una alta variabilidad dentro de cada muestra para cada variable estudiada, independientemente de la procedencia del genoma mitocondrial. Con el fin de comprobar si esa variabilidad sigue algún patrón que permita asociar las muestras con su origen celular o el género, hemos realizado un análisis de componentes principales (PCA) a través de QuasiFlow para estudiar un total de 23 variables de forma conjunta. Estas variables se distribuyeron en el espacio bidimensional del PCA como se indica en la **Figura 17**.

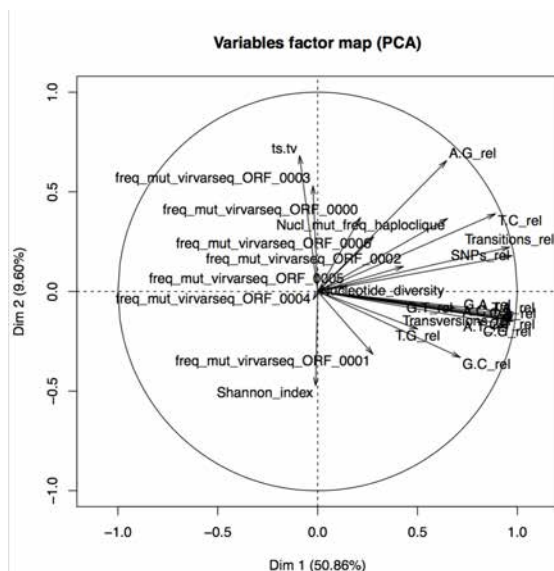


Figura 17. Mapa de distribución de las 23 variables que definen el espacio bidimensional de análisis de las muestras de DNA mitocondrial. Cada variable está representada por un vector con fuerza y dirección determinados que sitúan a cada muestra en una zona del mapa dependiendo de las características de estas variables. El eje X o dimensión 1 acumula el 50,86% de la variabilidad encontrada entre muestras, mientras que el eje Y o dimensión 2, acumula el 9,60% de dicha variabilidad.

Los resultados del análisis de componentes principales se muestran en la **Figura 18**. La mayoría de las muestras se ubican dentro de la misma zona del espacio bidimensional, lo cual indica cierto grado de homogeneidad entre las muestras. Sin embargo, QuasiFlow fue capaz de identificar un grupo de muestras con diferencias significativas con respecto al resto (representadas en color rojo y verde, **Figura 18**)

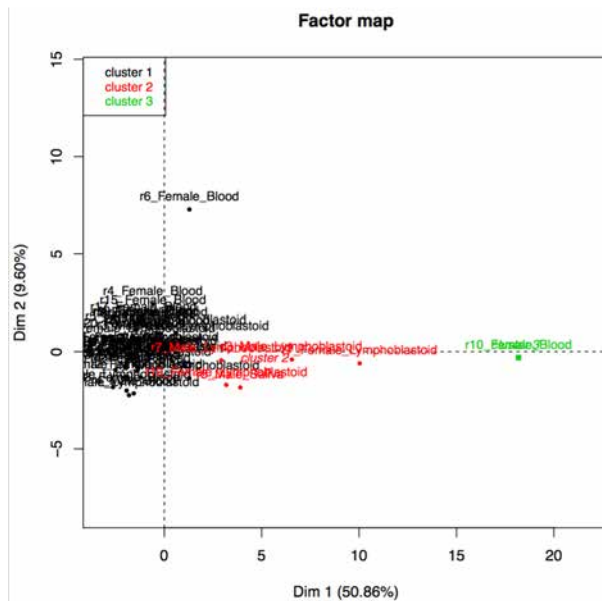


Figura 17. Mapa de distribución de las 23 variables que definen el espacio bidimensional de análisis de las muestras de DNA mitocondrial. Cada variable está representada por un vector con fuerza y dirección determinados que sitúan a cada muestra en una zona del mapa dependiendo de las características de estas variables. El eje X o dimensión 1 acumula el 50,86% de la variabilidad encontrada entre muestras, mientras que el eje Y o dimensión 2, acumula el 9,60% de dicha variabilidad.

Para averiguar si las diferencias entre las muestras analizadas mediante PCA indicadas en la **Figura 18** eran significativas, realizamos un cladograma con QuasiFlow (**Figura 19**). Esta función permite representar y agrupar por clados las muestras más semejantes entre sí, tomando como base las 23 variables analizadas por PCA. La **Figura 19** revela un claro agrupamiento entre muestras con DNAmT procedente de sangre, y un segundo clúster con una mayor proporción de muestras procedente de la línea linfoblastoide. En este estudio carecemos de muestras de distinto tipo celular (línea linfoblastoide, sangre y saliva) de un mismo individuo que nos permitiría comprobar si las observaciones realizadas son significativas. Sin embargo, los resultados evidencian la sensibilidad de QuasiFlow para detectar pequeñas diferencias entre muestras de DNA mitocondrial.

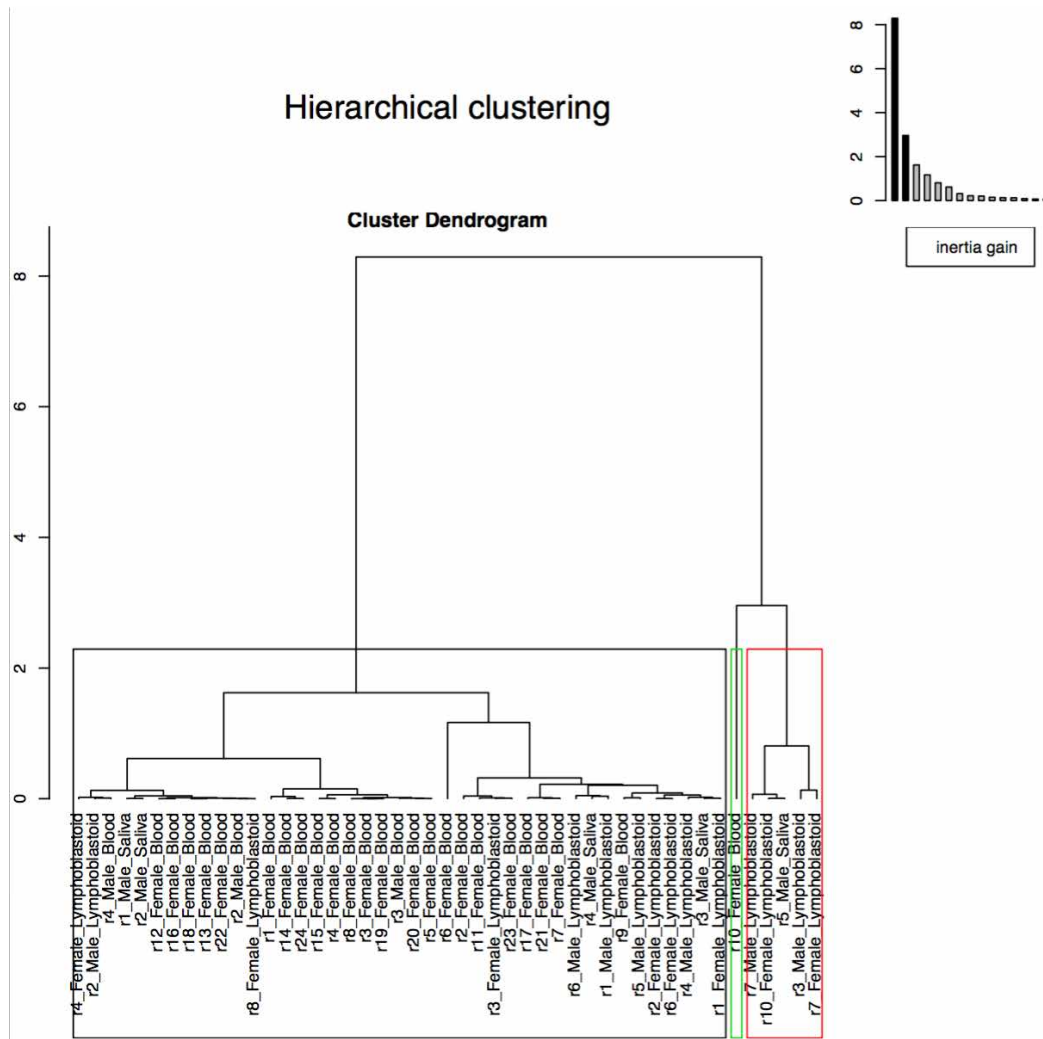


Figura 19. Dendrograma generado por QuasiFlow a partir del PCA. Las muestras analizadas se agrupan en clústeres en función de los resultados del PCA.

Discusión

La secuenciación masiva se está convirtiendo en una herramienta realmente útil a la hora de estudiar los cambios que suceden dentro de las poblaciones virales. Pero lleva consigo el hándicap de la dificultad que supone trabajar con la enorme cantidad de datos que genera cada secuenciación de estas características. La comunidad científica está empleando gran cantidad de tiempo y esfuerzo en desarrollar herramientas bioinformáticas capaces de analizar cada uno de los aspectos que caracteriza la complejidad de una cuasiespecie viral. No obstante, aún no se ha conseguido dar un salto cualitativo en este aspecto lo que ha causado que el uso de estas herramientas no termine de asentarse en la comunidad científica.

Recientemente, Gregori y colaboradores (Gregori et al, 2016) hacían una reflexión sobre el estado del arte en el estudio de las cuasiespecies virales a través de NGS. Estos autores remarcaban que, idealmente, el estudio de la complejidad de una cuasiespecie viral se debe basar en el conocimiento de la población completa y que dicho objetivo en este momento es inviable, aunque probablemente podrá alcanzarse dentro de pocos años.

En la actualidad, sin embargo, estamos trabajando con una metodología mucho menos perfecta, muy lejos del ideal, que obliga a una extrapolación a lo que podría ser la verdadera complejidad de toda la población a partir del análisis de una muestra. Además, como secuencia de la longitud de las lecturas, hasta hoy no era posible la reconstrucción completa de genomas virales, por lo que la mayoría de estudios se han limitado al análisis de algunas regiones de interés (amplicones).

En este contexto hemos presentado la herramienta QuasiFlow. Este flujo de trabajo es único en su enfoque. Se trata de un software centrado en el análisis multivariante y multimuestral. Es decir, basa su estudio en la robustez que ofrece el estudio de una gran cantidad de variables, que definen en su conjunto a cualquier cuasiespecie, y en su posterior uso conjunto para hacer análisis comparativos entre muestras. Dichas variables incluyen SNPs, InDels, eventos de recombinación, frecuencia de mutación, diversidad nucleotídica e índice de Shannon normalizado. Todo ello se lleva a cabo de forma automatizada y sencilla, sin necesidad de entrar en escritura de código. El punto fuerte de QuasiFlow es que permite la reconstrucción de cualquier cuasiespecie viral a

un nivel nunca antes visto, proporcionando un enfoque visual de la misma en el que pueden apreciarse relaciones de parentesco entre los diferentes haplotipos reconstruidos.

QuasiFlow es una herramienta de software libre creada con software de código abierto, que estará disponible de forma gratuita a través de la Plataforma de PICCASO, perteneciente al SCBI, en cuanto sea publicada.

Los resultados que arroja QuasiFlow van siempre acompañados de un robusto análisis estadístico, donde todas las variables son analizadas de forma individual para cada cuasiespecie, y comparadas dos a dos. De esta forma, el flujo de trabajo ofrece resultados que no necesitan posterior procesamiento creando gráficas de cada variable para cualquier número de muestras.

La otra gran ventaja de QuasiFlow es que permite ejecutar el análisis de todas las muestras de interés en paralelo, ahorrando una cantidad de tiempo apreciable. Además QuasiFlow ha sido desarrollado para poder diferenciar entre variantes reales dentro de una población y errores de secuenciación, reduciendo el error de la técnica hasta valores similares a los encontrados en la secuenciación Sanger (10^{-5} errores por nucleótido secuenciado). La herramienta también permite llevar un rastreo de las secuencias durante todo el proceso de análisis, creando tablas de seguimiento de lecturas gracias al número identificador de cada una de ellas lo que hace posible la trazabilidad y rescate de las secuencias analizadas en cualquier momento. Estos controles aseguran que los resultados sean veraces y comprobables.

La versatilidad del flujo ha sido puesto a prueba con gran éxito en el presente capítulo realizando búsquedas de recombinantes en coinfecciones de geminivirus, reconstruyendo la población viral y detectando haplotipos recombinantes dentro de ella de manera novedosa. Esto abre las puertas a una gran cantidad de estudios relacionados con la emergencia de virus que hasta ahora no era posible abordar. QuasiFlow ha sido probado con muestras de virus de RNA y de DNA, con secuenciaciones del genoma completo o amplicones, secuenciadas por Illumina HiSeq o Miseq (resultados no mostrados) demostrando la adaptabilidad del software a las necesidades del investigador .

También hemos demostrado la capacidad de QuasiFlow de detectar diferencias en genomas mitocondriales. Pese no ser el objetivo primario del flujo, la posibilidad de trabajar con cualquier número de muestras y de cualquier naturaleza, nos ha permitido obtener conclusiones muy interesantes para el estudio de la variabilidad genética de mitocondrias, un campo aún por explorar. A rasgos generales QuasiFlow ha podido identificar la naturaleza de los distintos orígenes de las muestras de DNAm_t, poniendo de relevancia la variabilidad existente dentro de los genomas mitocondriales.

Todos estos resultados invitan a pensar en QuasiFlow como el Software que la comunidad científica estaba esperando para dar el salto de calidad en el análisis de las cuasiespecies virales.

CAPITULO 3

**Efecto de las polimerasas de translesión y PrimPol en la
variabilidad genética del virus de ssDNA *Tomato yellow leaf
curl virus in vivo***

Introducción

1. Origen de la variabilidad genética de los virus de ssDNA

La mayor parte de los estudios de evolución de virus se han centrado en virus con genoma de RNA de animales ya que a este grupo de virus pertenecen la gran mayoría de los virus emergentes como el HIV, VHC y el virus Ébola entre otros. Sus características altas tasas de mutación y de sustitución nucleotídica (Drake, 1993; Hicks and Duffy, 2014; Sanjuán et al., 2010) son, entre otros factores, debidas a la baja fidelidad de las RNA y DNA polimerasas dependientes de RNA (RdRp y RdDp) que los replican, dado que carecen de actividad correctora de errores. En el caso de los virus de DNA, las bajas tasas de mutación que presentan (Sanjuán et al., 2010) se atribuyen a la mayor fidelidad de las polimerasas de DNA celulares implicadas en su replicación. Sin embargo, en virus de ssDNA emergentes como el anellovirus humano *Torque Teno virus* (Spandole et al., 2015), el parvovirus autónomo carnívoro o el virus adeno asociado de primate se ha observado un alto grado de heterogeneidad genética en parte dependiente de las relaciones particulares virus-huésped (López-Bueno et al., 2003; 2006) .

En virus emergentes de plantas también se ha observado una alta diversidad genética (apartado II de la Introducción General). Estos virus presentan estructuras poblacionales en cuasiespecies similares a las observadas para virus RNA animales. Recientemente se ha publicado que el poliomavirus JC, un virus de DNA circular de doble cadena (dsDNA) de humanos replicado por una DNA polimerasa celular desconocida, es altamente variable y está presente como cuasiespecies tanto en muestras de orina de individuos asintomáticos (Van Loy et al., 2013) como en oligodendrocitos de pacientes con leucoencefalopatía multifocal progresiva (Van Loy et al., 2015).

La frecuencia de mutación de un virus está determinada por una combinación de diferentes factores: I) Fidelidad de replicación de la polimerasa que lo replica; II) Capacidad de reparación de desapareamientos, III) Susceptibilidad a daños endógenos o exógenos producidos sobre el DNA como la oxidación, desaminación, producción de dímeros de timina por radiación UV, etc, IV) Desequilibrios en la proporción del *pool* de los diferentes nucleótidos, V) Deriva genética o la propia selección específica del virus resultante de los cuellos de botella a los que se ve sometida la población de virus, VI)

Acción de otras enzimas celulares que modifican las bases y VII) Bloqueos durante la replicación que pueden introducir mutaciones de tipo inserción o delección en los genomas virales o eventos de recombinación. La tasa de mutación vírica también pueden depender de la especie huésped que está infectada aunque los mecanismos responsables de esta variación no se conocen (Hicks and Duffy, 2014; Pita et al., 2007). Además, la tasa de mutación pueda variar a lo largo de la secuencia del genoma viral (Pathak and Temin, 1992), por ejemplo, a causa de la presencia de estructuras secundarias en el DNA. Así, se ha demostrado *in vitro* que las estructuras de tipo horquilla producen errores de replicación de tipo *slippage* que generan delecciones en el DNA (Viguera et al., 2001).

Las altas tasas de mutación de geminivirus podrían deberse a que las cadenas de ssDNA escapan a los mecanismos de reparación de desapareamientos (Duffy and Holmes, 2008; Ge et al., 2007; Roossinck, 1997) y, por otro lado, a mecanismos de desaminación espontánea del ssDNA o debido a la acción de enzimas desaminantes del huésped (Chelico et al., 2006; Stasolla et al., 2003) En este sentido, (Duffy and Holmes, 2008) analizaron secuencias publicadas de virus pertenecientes al complejo que produce la enfermedad del rizado amarillo del tomate (TYLCD) obtenidas los últimos 20 años y encontraron que las mutaciones fijadas presentaban un claro sesgo hacia cambios de tipo C→T y A→G, indicativas de un aumento en la tasa de desaminación de citosina a uracilo y de guanina a xantina (que empareja con timina), que contribuyen a aumentar por tanto el contenido en AT de los genomas virales.

De manera similar, en experimentos de evolución a corto plazo del mastrevirus MSV se ha encontrado un desequilibrio mutacional entre cadenas virales y complementarias del tipo G→T lo que apoya fuertemente la hipótesis de que las tasas de mutación elevadas de MSV serían consecuencia de un proceso mutagénico, posiblemente por oxidación de la Guanina hacia 8-oxoguanina (8-oxo-dG) (van der Walt et al., 2008).

Actualmente se desconoce qué DNA polimerasa celular de la planta replica los virus de ssDNA. De hecho, no se sabe si está implicada una única DNA polimerasa o bien si la replicación de virus ssDNA es un proceso que comprende diversas etapas y varias DNA polimerasas celulares. La replicación del DNA en plantas no ha sido completamente

dilucidada, si bien se cree que los principios básicos de replicación que se aplican a células animales también se aplican a plantas.

2. Polimerasas de DNA

Las DNA polimerasas juegan un papel concertado en la célula, asegurando tanto la replicación fiel del genoma durante cada división celular como su integridad. Tanto organismos procariotas como eucariotas poseen DNA polimerasas implicadas en replicación y DNA polimerasas especializadas en reparación de daño que puede causar el bloqueo de las DNA polimerasas replicativas. Así, en los genomas de los organismos eucariotas han sido identificadas DNA polimerasas replicativas nucleares (Pol α , δ y ϵ) y de orgánulos (Pol γ -like, Pol I-like en plástidos y Pol γ en mitocondrias), y numerosas DNA polimerasas de reparación (Pol β , λ , μ , σ , y TdT) y de tolerancia al daño o TLS (Pol θ , ν , ζ , η , ι , κ y Rev1) (**Figura 1**).

La frecuencia de mutación *in vivo* en eucariotas se estima en al menos 10^{-9} sustituciones por sitio (McCulloch and Kunkel, 2008). Este proceso viene asegurado por las tres DNA polimerasas replicativas Pol α , δ y ϵ (clasificadas dentro de la Familia A en función a su similitud de secuencia aminoacídica). Los análisis de fidelidad de replicación *in vitro* estiman un error por sustitución de base del orden de 10^{-5} a 10^{-6} . Además, otras DNA polimerasas tienen un papel específico en la replicación del genoma nuclear: Pol ζ (Familia B) y los miembros de la Familia Y (Pol η , ι , κ y Rev1), que realizan la síntesis de translesión (TLS) (Lehmann et al., 2007; Waters et al., 2009), junto con las DNA polimerasas Pol β , λ y μ (Familia X) o Pol ν y θ (Familia A). La fidelidad de replicación de miembros de la Familia X es del orden de 10^{-3} a 10^{-4} , mientras que la fidelidad más reducida corresponde a miembros de la Familia Y (10^{-1} a 10^{-3}) (**Figura 2**).

En las plantas, en cambio, no se han identificado algunas de las polimerasas descritas en animales o levaduras como Pol γ , Pol ϕ , Pol β , Pol μ y TdT (Burgers et al., 2001; Garcia-Diaz and Bebenek, 2007; Kimura and Sakaguchi, 2006). Hasta la fecha se piensa que existen unas 11 polimerasas de DNA en *A. thaliana* cuya función, para la mayor parte de ellas, no está bien definida (Garcia-Diaz and Bebenek, 2007).

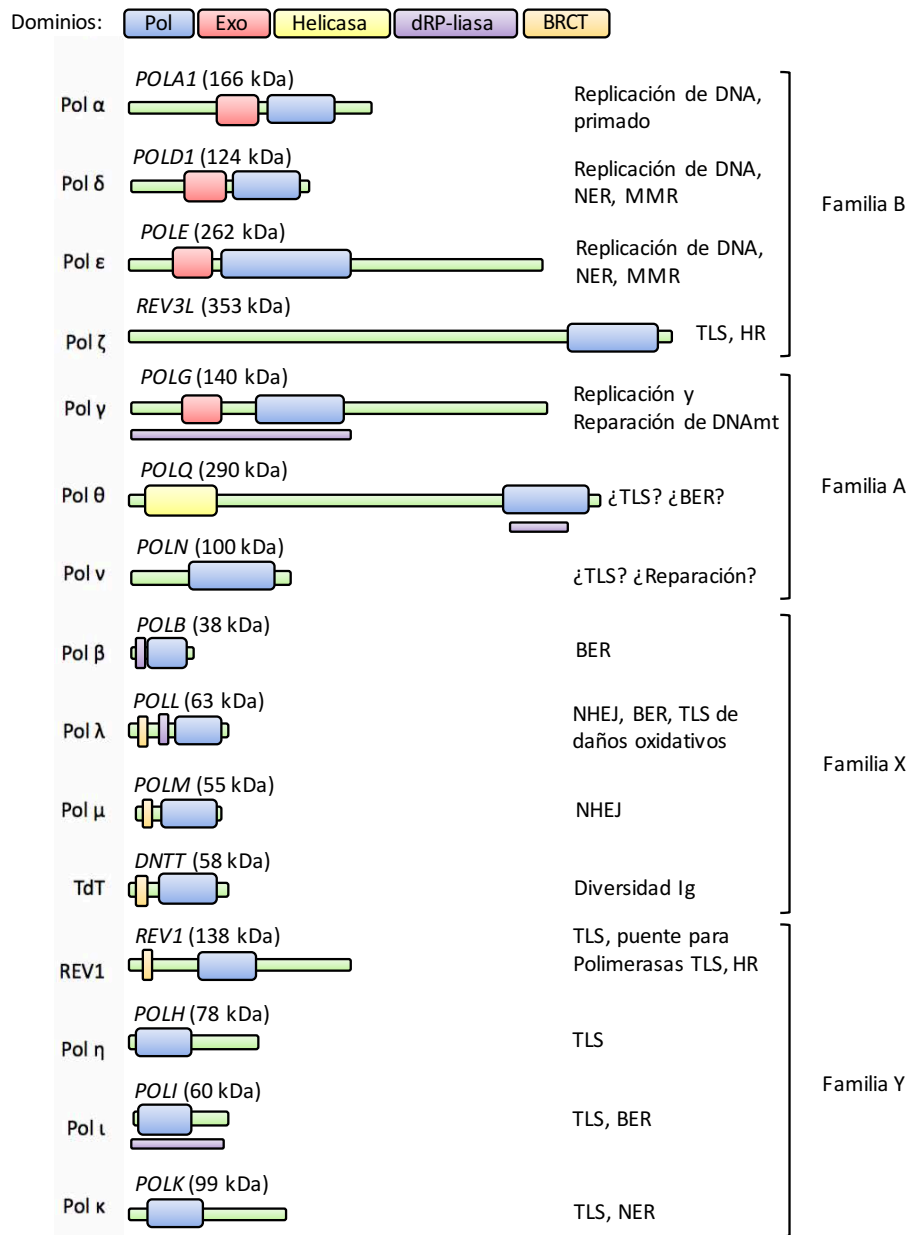


Figura 1. DNA polimerasas humanas dependientes de DNA. Hasta la fecha han sido caracterizadas 16 DNA polimerasas dependientes de DNA en humanos, cada una de ellas nombrada mediante letras griegas. En la figura se muestran los dominios que forman parte de la subunidad catalítica de estas Polimerasas. Azul: dominio polimerasa; rojo: dominio 3'-5' exonucleasa; amarillo: dominio helicasa; morado: dominio 5'-desoxiribosa fosfato liasa; naranja: dominio C-terminal BCRA1 (BCRT). Aunque Pol α contiene un dominio 3'-5' exonucleasa, este es inactivo. Sobre cada polimerasa se ha situado el nombre del gen y el peso molecular de la subunidad catalítica. BER: reparación por escisión de base; NER: reparación por escisión de nucleótido; MMR: reparación de desapareamientos; TLS: síntesis de translesión; NHEJ: recombinación de extremos no homólogos; HR: recombinación homóloga. En plantas han sido descritas todas ellas excepto Pol γ, Pol φ, Pol β, Pol μ y TdT. Figura adaptada de Lange *et al.*, 2011.

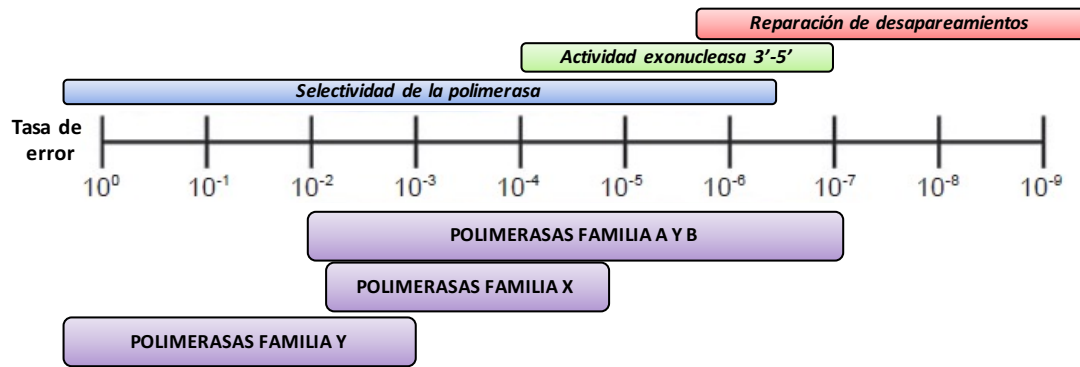


Figura 2. Fidelidad de la replicación de las diferentes familias de DNA polimerasas. En la parte superior de la figura se muestran las funciones responsables de la fidelidad de copia de las polimerasas estimadas a partir de estudios de mutagénesis. Debajo, la barra horizontal muestra los rangos de la tasa de error determinada *in vitro* para las diferentes familias de polimerasas y para el complejo de replicación completo *in vivo*. Dentro de cada familia, las tasas de error pueden variar ampliamente entre polimerasas y tipos de errores. Adaptación de Duffy and Holmes, 2008.

Estudios cristalográficos han revelado que las subunidades catalíticas de todas las DNA polimerasas comparten un dominio de polimerización constituido por una misma cadena polipeptídica plegada formando tres subdominios denominados *palm* (palma), *thumb* (pulgar) y *fingers* (dedos) por su similitud con la forma de una mano derecha parcialmente abierta (**Figura 3**). El subdominio *palm*, posee el centro catalítico para la reacción de polimerización que facilita la unión del extremo 3'OH de la molécula iniciadora con el grupo α -fosfato del nucleótido complementario al de la cadena molde. El subdominio *dedos* permite la unión y orientación de la cadena molde y es responsable de la unión del nucleótido entrante a la DNA polimerasa. Por último, el subdominio *pulgar*, proporciona estabilidad al complejo DNA polimerasa/DNA por su unión al surco

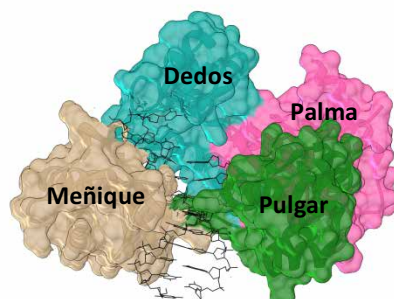


Figura 3. Dominios dentro de la subunidad catalítica de Pol η. Estructura cristalográfica de Pol η humana, cuyo homólogo en *A. thaliana* es AtPolH, presentando la clásica estructura de mano derecha. Pueden observarse los dominios más importantes del sitio catalítico de las TLS: los dedos (azul), la palma (rosa), el pulgar (verde) y el meñique (marrón). En la figura el pulgar, el meñique y los dedos abrazan la doble hebra de DNA.

menor del dúplex de DNA en síntesis y permite la entrada del complejo estructura molde/iniciador al centro activo.

3. Síntesis de translesión en eucariotas

La síntesis de translesión es un mecanismo universal por el cual las células pueden tolerar cierto nivel de daño producido en el DNA. Está basado en la propiedad de ciertas DNA polimerasas especializadas en reparar un DNA molde que contiene una lesión que no puede ser replicada de forma normal por una DNA polimerasa replicativa, posibilitando completar la duplicación del genoma. En este proceso puede intervenir una o varias polimerasas TLS, en cuyo caso una polimerasa inserta un nucleótido frente a la lesión y una segunda polimerasa extiende el extremo 3' OH, hasta que la polimerasa replicativa pueda reanudar su acción (**Figura 4**). El grupo más abundante de polimerasas

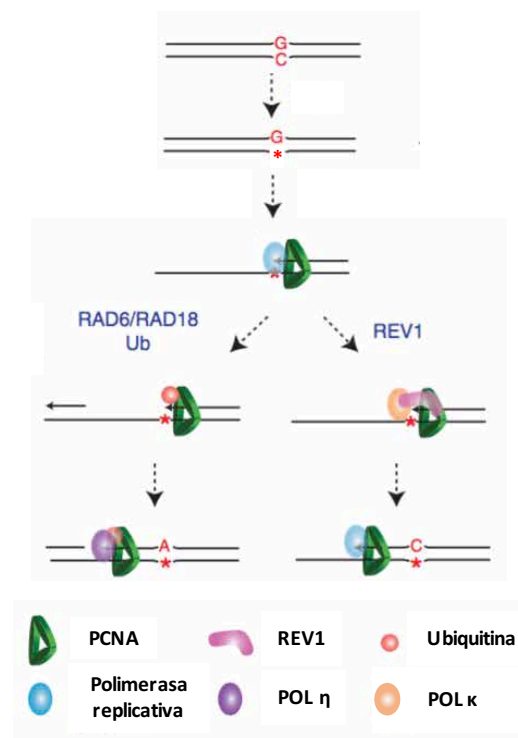


Figura 4. Reparación de DNA mediada por síntesis de translesión (TLS). Cuando la maquinaria replicativa se bloquea debido a un daño en el DNA molde se activa una cascada de señales para la activación de TLS. Estas pueden actuar mediante dos mecanismos: en el primero RAD18 ubiquitina PCNA y esto permite la unión de Pol η que extenderá la cadena hasta el lugar de reinicio de la replicación tras el daño en la cadena molde. En el segundo mecanismo, Rev1 se une a PCNA sin necesidad de que este se ubiquitine haciendo de puente de unión entre PCNA y otras TLS como Pol κ , que permite la replicación del daño producido en la cadena molde.

especializadas en TLS son las pertenecientes a la Familia Y, que en eucariotas superiores se compone de Pol η , ι , κ , y REV1, aunque en plantas no se ha identificado Pol ι . Además, varias de las polimerasas de la Familia X y la polimerasa ζ de la Familia B también están implicadas en procesos de TLS. Recientemente se ha descubierto una primasa-polimerasa denominada PrimPol que también podría, entre otras muchas funciones, estar implicada en procesos de TLS (García-Gómez et al., 2013; Rudd et al., 2014).

3.1. Familia Y

Las DNA polimerasas de la Familia Y se caracterizan por su capacidad para replicar a través de bases dañadas, una baja procesividad y una fidelidad reducida debido a que carecen de actividad de corrección de errores. Estas polimerasas poseen un centro activo mucho más amplio que el de las DNA polimerasas replicativas de la Familia A y B lo que permite aceptar estructuras de DNA voluminosas. Otra diferencia en los miembros de la Familia Y con respecto a las polimerasas replicativas es el menor tamaño de los subdominios “dedos” y “pulgar”, lo que permite un menor contacto con la molécula de DNA y con el nucleótido entrante (**Figura 5**). Esto facilita la actividad TLS y también contribuye a una baja fidelidad de copia.

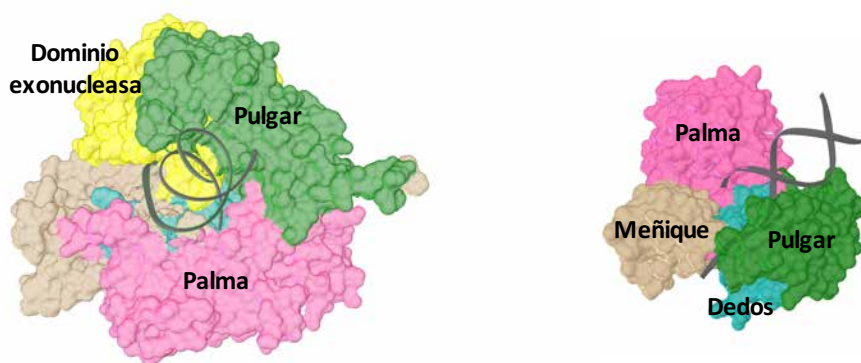


Figura 5. Comparativa entre los dominios dentro de la subunidad catalítica de una polimerasa replicativa y una polimerasa TLS. En la figura pueden observarse las diferencias de tamaño entre los dominios más importantes del sitio catalítico de las polimerasas replicativas y de las polimerasas TLS. Los dedos (azul), la palma (rosa), el pulgar (verde) y el meñique (marrón).

Además, las polimerasas de la Familia Y carecen de un dominio de corrección 3'-5' exonucleasa, contribuyendo aún más a la baja fidelidad de síntesis. Estas polimerasas también poseen un sub-dominio adicional, denominado dominio asociado a polimerasas

(PAD) (Trincao et al., 2001) o "dedo meñique" (Ling et al., 2001), el cual es el menos conservado entre las polimerasas de la Familia Y y está implicado en la especificidad de cada polimerasa hacia un tipo de lesión en concreto (Boudsocq et al., 2004).

4. DNA polimerasas TLS descritas en plantas y animales

4.1. DNA polimerasa η

Pol η (también llamado Rad30A) es quizás la polimerasa mejor caracterizada de la Familia Y. Ha servido como modelo para los demás miembros de esta Familia, debido en gran parte a la relevancia clínica de Pol η , donde el gen que codifica para esta polimerasa aparece mutado en la enfermedad XP-V (Xeroderma pigmentosum variant) (Johnson et al., 1999; Lehmann et al., 2011; Masutani et al., 1999). Se ha demostrado la implicación de esta polimerasa en la reparación del daño en el DNA causado por UV, pudiendo replicar dímeros de timina (Johnson et al., 1999; Masutani et al., 1999) con una eficiencia similar a la replicación del DNA no dañado (Biertümpfel et al., 2010a; McCulloch and Kunkel, 2008; Silverstein et al., 2010). El centro activo de Pol η es amplio, de tal forma que es capaz de acomodar ambas timinas que conforman el dímero, estabilizándolas para permitir el emparejamiento correcto con adenina. Pol η puede también reparar otro tipo de lesiones *in vitro* como productos 8-oxo-G (revisado en Haracska et al., 2000; Lehmann, 2002; Zhang et al., 2000), aunque de forma menos eficiente que los dímeros de timina y no parece ser una función *in vivo* de esta enzima (Avkin and Livneh, 2002). Pol η tiene una baja fidelidad de copia de DNA no dañado (Matsuda et al., 2000) y su acción está limitada debido a su baja procesividad (Biertümpfel et al., 2010b). Pol η también parece estar implicada en otros procesos que no requieren actividad TLS como recombinación homóloga (McIlwraith et al., 2005), replicación del DNA no dañado durante la fase S, posiblemente en la replicación de los sitios frágiles de cromosomas (Rey et al., 2009), la hipermutación génica durante el desarrollo del sistema inmune (Zeng et al., 2001), y la reparación de agrupamientos de daños oxidativos en el DNA (Zlatanou et al., 2011).

En *A. thaliana*, la polimerasa homóloga a Pol η fue descrita por primera vez por (Kunz et al., 2005) como parte de las polimerasas celulares implicadas en la reparación por escisión de nucleótidos (NER). También se ha descrito su implicación en la reparación al daño por UV (Anderson et al., 2008; Jesús Santiago et al., 2008; Nakagawa et al., 2011a;

2011b) y su interacción con PCNA2 (Kunz, 2008; Masuda et al., 2015)

4.2. DNA polymerase κ

Pol κ (también llamada DinB) se identificó en la búsqueda de homólogos eucariotas para el gen de *E. coli* DinB (Gerlach et al., 1999; Johnson et al., 2000; Ogi et al., 1999), codificante de la polimerasa de TLS Pol IV de *E. coli*, siendo la única polimerasa de la Familia Y que se conserva en todos los filos en los que se ha estudiado (Ohmori et al., 2001). Como todas las polimerasas de la Familia Y se trata de una DNA polimerasa de baja fidelidad, aunque es la más precisa de esta familia (Johnson et al., 2000) debido a que el sitio activo es relativamente restrictivo en comparación con otros miembros de la Familia Y; por ello, Pol κ es incapaz de reparar daños producidos por UV y tiene una baja eficiencia de TLS frente a lesiones *in vitro* (Ohashi et al., 2000; Zhang et al., 2000). Sin embargo, Pol κ extiende muy eficientemente pares de bases erróneos (Haracska et al., 2002; Washington et al., 2002), y esta función extensora de Pol κ es particularmente evidente en las células XP-V (Ziv et al., 2009).

En *A. thaliana* también se ha descrito una polimerasa homóloga llamada PolK (García-Ortiz et al., 2004). La función de esta polimerasa en plantas es similar a la que lleva a cabo su homóloga en animales. Se ha demostrado su implicación en reparación NER (Kunz et al., 2005), y más recientemente se ha demostrado su presencia en todos los tejidos de la planta (Santiago et al., 2008) remarcando la importancia de esta polimerasa dentro de la célula vegetal.

4.3. Rev1

Rev1 no es estrictamente una DNA polimerasa si no una dCMP transferasa, capaz de incorporar dCMP en cadenas molde con guaninas (Nelson et al., 1996). La capacidad como TLS de Rev1 es limitada, pudiendo incorporar de manera eficiente citosinas a estructuras voluminosas con alto contenido en guaninas (Zhang et al., 2002) y en sitios abásicos (Nelson et al., 1996). Se ha demostrado que el *bypass* de sitios abásicos es una función *in vivo* de Rev1, ya que mutantes catalíticos de Rev1 alteran el espectro de mutaciones en linfocitos B de vertebrados (Jansen et al., 2006; Masuda et al., 2009; Ross and Sale, 2006). La estructura cristalina de Rev1 reveló la base molecular de su función celular: en lugar de utilizar la clásica unión Watson-Crick para detectar el emparejamiento correcto de la dCMP entrante con el DNA molde, los residuos en el

“dedo meñique” de Rev1 coordinan temporalmente la orientación de la cadena molde girando una de las base de esta cadena hacia fuera, mientras que la dCMP se une mediante puentes de hidrógeno a un residuo de arginina. La palma del dominio catalítico de Rev1 crea espacio para acomodar esta estructura y los puentes de hidrógeno formados por la dCMP entrante y el residuo de arginina elimina la necesidad de una base de guanina para la incorporación correcta de la citosina (Nair et al., 2005). REV1 está implicada en la mutagénesis inducida por UV en levaduras (Lemontt, 1971) y en humanos (Gibbs et al., 2000). Aunque curiosamente, no se requiere la actividad catalítica de Rev1 para la mutagénesis de linfocitos B, ya que la mutagénesis se sigue produciendo aún en un mutante catalítico de Rev1, pero el espectro de mutaciones se altera (Otsuka et al., 2005).

El extremo carboxilo-terminal de Rev1 interactúa con las otras polimerasas Y-Familia y Pol ζ (Guo et al., 2003; Murakumo et al., 2001; Ohashi et al., 2004; Tissier et al., 2004), mientras que su extremo amino- terminal contiene un dominio BRCT (BRCA1 carboxilo-terminal) que es importante para la interacción con PCNA y esencial para la supervivencia tras exposición a la luz UV (Guo et al., 2006). Estos resultados muestran el papel esencial de REV1 como puente en la interacción de PCNA con el resto de polimerasas TLS, siendo un regulador clave en TLS. Recientemente se ha atribuido a REV1 la replicación a través de estructuras terciarias de DNA como los G-cuadrupeles con grandes implicaciones en la regulación génica (Perrone et al., 2014; Sarkies et al., 2010; van Kregten and Tijsterman, 2014), y recombinación homóloga (HR) (Sharma et al., 2012).

En plantas también ha sido identificada el homólogo de Rev1. Así, en *A. thaliana* AtRev1 parece estar implicada en la reparación de daño por UV (Nakagawa et al., 2011b; Takahashi et al., 2005) y tolerancia al daño oxidativo (Amoroso et al., 2011; Takahashi et al., 2007).

4.4. PrimPol una nueva primasa polimerasa implicada en TLS

Recientemente ha sido descrita en humanos una nueva primasa-polimerasa perteneciente a la familia AEP (primasas de arqueas y eucariotas), denominada PrimPol (García-Gómez et al., 2013). Esta enzima tiene tanto actividad primasa (RNA y DNA primasa) como DNA polimerasa, extendiendo un cebador pre-existente utilizando

preferentemente dNTPs. PrimPol se detecta tanto en mitocondrias como en el núcleo celular. PrimPol es capaz de sintetizar un cebador de DNA o de RNA ya que tiene la misma afinidad por desoxirribonucleótidos que por ribonucleótidos, pudiendo incorporar cualquiera de éstos para la síntesis de una nueva cadena. En el caso de que ya exista algún nucleótido incorporado, PrimPol puede continuar la extensión de la cadena incorporando desoxirribonucleótidos, debido a su mayor afinidad por éstos cuando dispone de un extremo 3' libre (García-Gómez et al., 2013). PrimPol carece de actividad correctora de errores, con una fidelidad de copia similar a la de la familia X ($5,1 \times 10^{-4}$) aunque con una tasa de introducción de deleciones similar a las DNA polimerasas de la familia Y (Martínez-Jiménez et al., 2015). PrimPol es altamente procesiva y se ha demostrado que es capaz de actuar como TLS, replicando a través de lesiones en el DNA como fotoproductos de UV (Bianchi et al., 2013), especialmente los 8-oxo-dG (García-Gómez et al., 2013) (**Figura 6**). También se ha demostrado que puede primar el DNA aguas abajo de una lesión, lo que facilitaría la posterior reincorporación de la polimerasa replicativa y además, actúa en sitios abásicos introduciendo C o A, teniendo mucha más afinidad por la introducción de C (García-Gómez et al., 2013). Por otro lado, hay indicios de que podría estar implicada en procesos de recombinación (Martínez-Jiménez et al., 2015). PrimPol es independiente de unión a PCNA y su actividad es inhibida por RPA (Guilliam et al., 2015).

Debido a su plasticidad y su acción como TLS, en esta Tesis hemos estudiado el posible papel de PrimPol en procesos de reparación de daño en virus de ssDNA como los geminivirus, así como en el inicio de la replicación en ausencia de DNA truncado que dé inicio a la RDR, gracias a su actividad primasa.

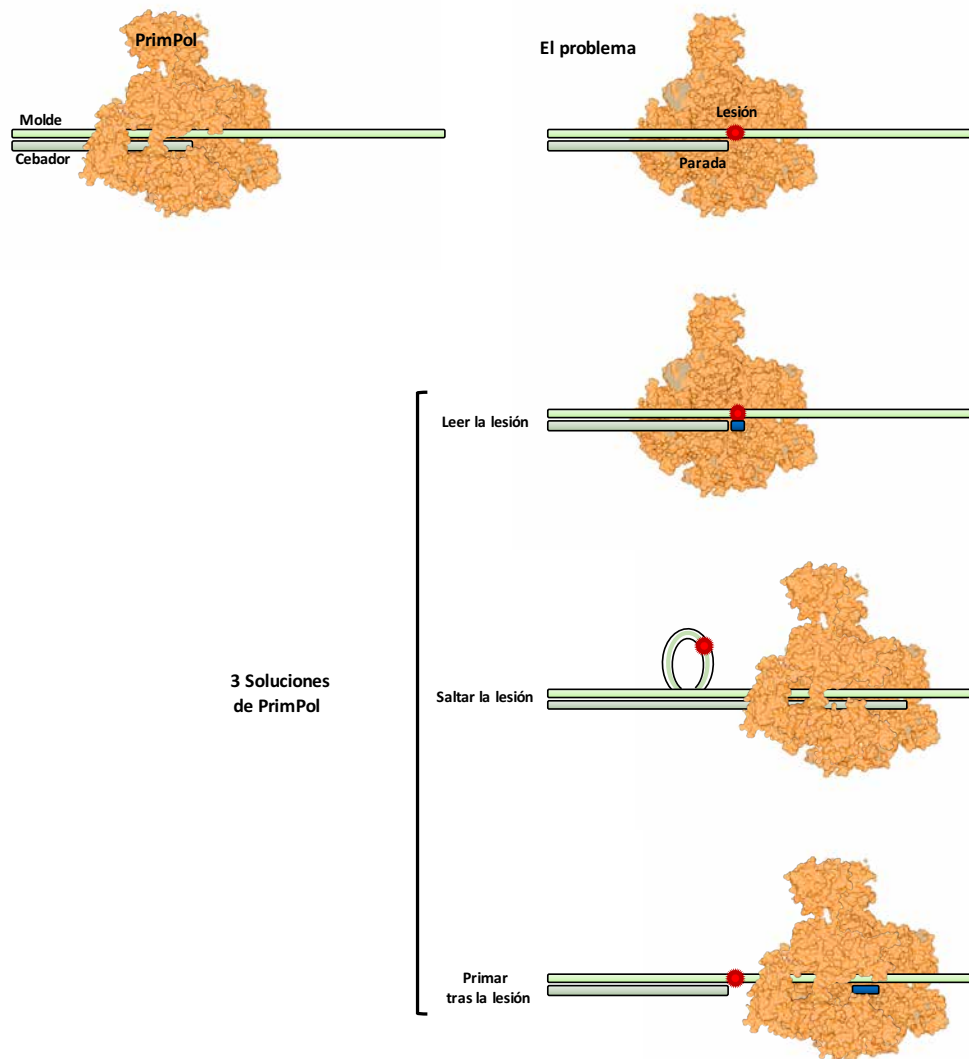


Figura 6. Soluciones alternativas para la reparación de DNA mediada por la actividad TLS de PrimPol humana. La replicación del DNA se detiene cuando la polimerasa replicativa se encuentra con una lesión en el ADN molde, acumulando ssDNA delante de la lesión. PrimPol reconoce este tipo de daño y es capaz de resolver el bloqueo replicativo por 3 mecanismos diferentes, dependiendo de la naturaleza de la lesión: (1) lesiones tipo 8-oxo-dG, las corrige directamente actuando como una TLS polimerasa convencional; (2) lesiones ilegibles (sitios abásicos, 6-4PP, etc.), es capaz de saltar por microhomología mediada por la realineación antes de la lesión, actuando como una "pseudo" TLS; (3) puede realizar la síntesis de un nuevo primado tras la lesión, como una primasa TLS. Adaptado de Martínez-Jiménez et al., 2015. 44

En *A. thaliana* existe un gen homólogo al gen primpol humano, el At5G52800, situado en el cromosoma 5. Este gen está anotado como una primasa, pero aún no existe ningún estudio que estudie su función, aunque el hecho de que se exprese diferencialmente en células acompañantes del floema, la hace un candidato interesante ya que los geminivirus sólo replican en estas células (Morilla et al., 2006).

5. TYLCV y el género *Begomovirus*

Actualmente, el ICTV reconoce 7 géneros en la familia *Geminiviridae* en función de su organización del genoma, su insecto vector y su gama de huéspedes: Becurtovirus, Begomovirus, Curtovirus, Eragrovirus, Mastrevirus, Topocuvirus y Turncurtovirus (Brown et al., 2015; Varsani et al., 2014) (**Figura 7**).



Figura 7. Organización genómica de los 7 géneros pertenecientes a la familia *Geminiviridae*. Las regiones del mismo color representan ORFs que codifican proteínas homólogas. Dentro del género *Begomovirus* se indica la distribución en los genomas monopartitos y en los bipartitos. Adaptado de Varsani et al., 2014.

Los géneros *Becurtovirus*, *Eragrovirus*, *Turncurtovirus* y *Topocurtovirus* son de reciente descubrimiento y cada uno de ellos incluye una o dos especies. Por su parte, los géneros *Begomovirus*, *Mastrevirus* y *Curtovirus*, con 288, 29 y 3 especies, respectivamente, son

los más importantes y más estudiados de todos los que componen la familia *Geminiviridae*. Además, desde la ICTV se proponen varios criterios de demarcación de especies de begomovirus, el más importante de ellos es el umbral de identidad nucleotídica del 91 % de la secuencia completa entre genomas A (Brown et al., 2015).

El género *Begomovirus* infecta plantas dicotiledóneas de forma persistente circulativa mediante la transmisión por mosca blanca. La mayoría de los miembros de este género poseen un genoma bipartito, cuyos componentes se denominan DNA-A y DNA-B. Los begomovirus se pueden clasificar en dos grupos filogenéticos: i) begomovirus del “Viejo Mundo”, que incluye especies procedentes de la cuenca mediterránea, África, Oriente Medio, Asia y Australia; ii) begomovirus del “Nuevo Mundo” procedentes del continente americano (Stanley, 2005). Algunas especies incluidas en el grupo del “Viejo Mundo” presentan genomas monopartitos, como la mayoría de los virus asociados con la enfermedad del rizado amarillo del tomate (TYLCD). Se han descrito 11 especies asociadas a TYLCD, siendo 9 de ellas de genoma monopartito, y sólo 2 de genoma bipartito. Las cepas de las especies de begomovirus están definidas por un umbral de identidad nucleotídica del 93%. Los síntomas causados por estas especies en tomate son muy similares entre sí, aunque algunas presentan diferencias de afinidad por biotipos del vector (Sánchez-Campos et al., 1999) o gama de huéspedes (García-Andrés et al., 2006).

La especie asociada a TYLCD que se encuentra más ampliamente distribuida por el mundo es *Tomato yellow leaf curl virus* (TYLCV) (Abhary et al., 2007; Lefeuvre et al., 2010). En España, la primera especie que se vio asociada a la enfermedad, *Tomato yellow leaf curl Sardinia virus* (TYLCSV), fue desplazada tras la entrada de TYLCV-Mld hasta minimizar su incidencia (Sánchez-Campos et al., 1999), en parte debido al uso extensivo de cultivares tolerantes (García-Andrés et al., 2009). En 2003, se describió la entrada de TYLCV-IL en España (Morilla et al., 2007), y dos nuevas especies en el sudeste peninsular resultantes de procesos de recombinación, muy frecuentes en este grupo (Lefeuvre et al., 2007; Martin and Rybicki, 2000), *Tomato yellow leaf curl Malaga virus* (TYLCMaV) (recombinante entre TYLCSV y TYLCV-Mld) y *Tomato yellow leaf curl Axarquía virus* (TYLCAxV) (entre TYLCSV y TYLCV-IL) (García-Andrés et al., 2006; Monci et al., 2002a). En áreas colonizadas por una sola especie se ha descrito una gran

estabilidad genética poblacional (Delatte et al., 2007; Sánchez-Campos et al., 1999). Sin embargo, cuando coexisten distintos tipos virales en un área se dan con frecuencia casos de infecciones mixtas, lo cual suele favorecer la aparición de nuevos tipos virales de naturaleza recombinante (García-Andrés et al., 2007). Estos nuevos genotipos recombinantes pueden llegar a desplazar a los tipos parentales, como se ha descrito para TYLCMaV y TYLCAxV (García-Andrés et al., 2007; Monci et al., 2002a).

5.1. Estructura y función de la región intergénica no codificante

En la región IR se ubican los elementos clave para la replicación y transcripción virales. En la región correspondiente al lazo de la horquilla se localiza el nonanucleótido TAATATT↓AC que está altamente conservado y sobre el cual la proteína iniciadora Rep produce una escisión (↓) para el inicio de la replicación (Gutierrez, 1999). La zona del tallo de la horquilla está formada por 46 nt (en los begomovirus monopartitos) y es rica en GC. En su extremo 3', la IR contiene una secuencia de unión específica de la proteína TrAP que transcribe los genes V1 y V2 (Ruiz-Medrano et al., 1999) y la caja TATA del promotor correspondiente. En su extremo 5', está la secuencia de unión de la Rep, constituida por motivos repetidos en tándem (iterones) cuya organización y secuencia

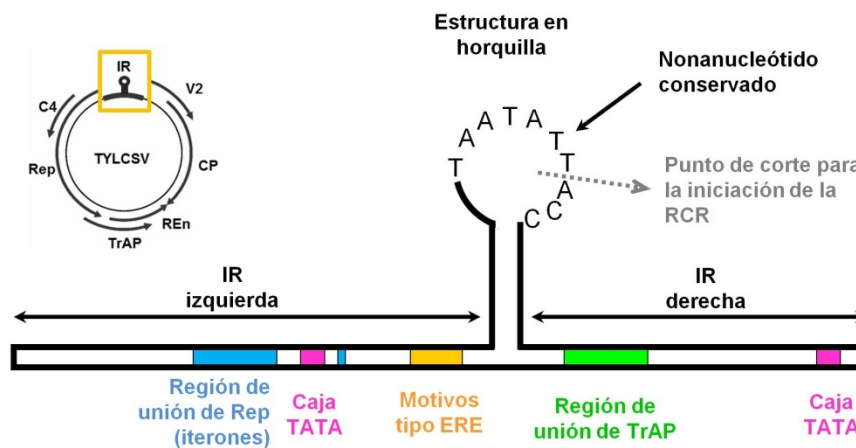


Figura 8. Representación esquemática de la región intergénica de los Begomovirus. En la estructura en horquilla se encuentra el nonanucleótido conservado que utiliza REP como punto de corte y liberación del monómero. A la derecha de la horquilla se halla la secuencia de unión de TrAP al DNA y la caja TATA para la expresión de los genes en sentido viral. En la parte izquierda están representadas la región genómica de unión de REP y la caja TATA para la expresión de genes en sentido antiviral. Tomado de la Tesis de Guillermo Dominguez Huerta, 2015.

nucleotídica dependen del linaje viral (Argüello-Astorga et al., 1994), todas estas características pueden observarse en la **Figura 8**.

5.2. Estructura y función de las proteínas virales

El genoma viral de los begomovirus asociados a TYLCD contiene 6 ORFs que codifican para 6 proteínas, 2 en sentido viral y 4 en sentido antiviral. Todos los ORFs presentan algún grado de solapamiento entre sí y están organizadas en dos unidades de transcripción inversas y separadas por la región intergénica (IR) de unos 300 nt de longitud (Martin and Rybicki, 2000; Moriones and Navas-Castillo, 2000; Picó et al., 1996).

Los ORFs en sentido viral son:

- V2: codifica para una proteína de 13,3 KDa que parece estar implicada en el movimiento a corta y a larga distancia dentro del hospedador (Rojas et al., 2001), la inducción de síntomas de la infección. (Wartig et al., 1997) y la supresión del silenciamiento génico en TYLCV (Zrachya et al., 2007). Se ha propuesto que V2 suprime el silenciamiento postranscripcional (PTGS) interfiriendo con la amplificación del siRNA, tanto por interacción directa con SGS3 como por competición por el dsRNA que produce la maquinaria de silenciamiento (Fukunaga and Doudna, 2009). Además, en tomate, la proteína V2 de TYLCV contrarresta las defensas de la planta inhibiendo la actividad de la proteína CYP1, una cisteín-proteasa miembro de las PLCP (*Papain-like cysteine proteases*), implicadas en la ruta de muerte celular programada, por lo que podría suprimir la respuesta hipersensible (HR) de la célula huésped (Bar-Ziv et al., 2012; 2015).
- V1/CP: codifica para una proteína de 30 KDa, única responsable de la encapsidación del genoma viral. Se ha demostrado que CP es necesaria para la transmisión y la especificidad de vector (Höhnle et al., 2001) y que protege el ssDNA genómico de la degradación durante la transmisión (Azzam and Frazer, 1994). Esta proteína es fundamental para la entrada y salida de los geminivirus en el núcleo, ya que posee señales de localización y exportación nuclear (NLS y NES) (Gafni, 2003). En el caso de begomovirus monopartitos, es indispensable para los movimientos del virus a corta y a larga distancia en la planta (Jupin et al., 1994; Noris et al., 1998). Asimismo, interviene en la replicación por interacción con Rep (Malik et al., 2005).

Las proteínas codificadas por los ORFs en sentido antiviral son las siguientes:

- C1/Rep (41 kDa): es la única proteína indispensable para la acumulación del DNA viral. Rep es una endonucleasa-helicasa (Laufs et al., 1995a) que reconoce específicamente el origen de replicación del virus e hidroliza y liga la cadena viral dentro del nonanucleótido conservado (Gladfelter et al., 1997; Orozco and Hanley-Bowdoin, 1996). También se ha descrito su actividad helicasa (Clérot and Bernardi, 2006) y su interacción con distintas proteínas de la célula huésped, tales como PCNA (*Proliferating Cell Nuclear Antigen*) y homólogos a la familia retinoblastoma (pRBR) (Ach et al., 1997; Castillo et al., 2003; Gutierrez, 1999; Luque et al., 2002). Se ha propuesto que Rep interfiere con la defensa epigenética del huésped, ya que disminuye el nivel de expresión de metiltransferasas de DNA implicadas en el silenciamiento transcripcional (TGS) (Rodríguez-Negrete et al., 2013).
- C4 (11 kDa): proteína a la que se atribuyen funciones de movimiento a corta y larga distancia y la expresión de síntomas (Jupin et al., 1994; Rojas et al., 2001; Viral Genome Methylation as an Epigenetic Defense against Geminiviruses, 2008). Además, es un supresor del PTGS, lo que permite un incremento de la invasión viral y de la inducción de los síntomas (Bisaro, 2006; Vanitharani et al., 2004). Su eficacia como supresor del PTGS es muy variable debido a que existe una gran diversidad de secuencias de esta proteína, a pesar de estar codificada por un gen completamente solapado con un gen tan conservado como el que codifica la proteína Rep (Bisaro, 2006; Vanitharani et al., 2004). Se ha demostrado que C4 actúa potenciando la actividad supresora del TGS ejercida por Rep (Rodríguez-Negrete et al., 2013).
- C2/TrAP (15,7 kDa): la proteína activadora de la transcripción (Transcription Activation Protein) es necesaria en la transcripción eficiente de los genes en el sentido viral. Se considera un factor de patogenicidad debido a su implicación en el movimiento a larga distancia (Wartig et al., 1997) y la supresión del PTGS (Dong et al., 2003; Vanitharani et al., 2004; Voinnet et al., 1999; Wang and Wagner, 2003). Además, interfiere con el TGS mediante la inactivación de la adenosin kinasa (ADK), que se requiere para la síntesis de S-adenosil metionina, un cofactor esencial para las metiltransferasas de DNA (Buchmann et al., 2009).
- C3/REn (16 kDa): la proteína potenciadora de la replicación (Replication en

hancer

) no es esencial para el ciclo de vida de los begomovirus, aunque contribuye al

aumento de la eficiencia del proceso de replicación, incrementando la acumulación de DNA viral y la expresión de síntomas (Hormuzdi and Bisaro, 1995; Selth et al., 2004; Sunter et al., 1990) mediante la interacción con Rep y con proteínas de la célula huésped (Castillo et al., 2003; Selth et al., 2005; Settlage et al., 2005).

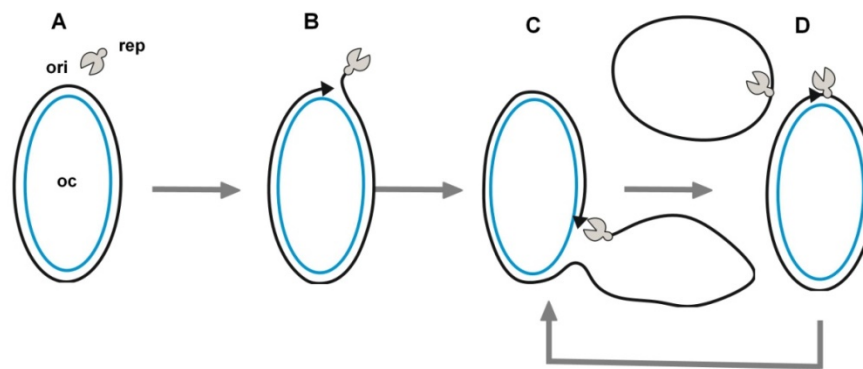
5.3. Estructura del virión y ciclo de vida

Las partículas virales de TYLCV presentan una morfología geminada característica de unos 30 nm de longitud y 18-20 nm de anchura, formada por 22 capsómeros pentaméricos organizados en dos icosaedros incompletos unidos por una de sus caras. En infecciones naturales, el momento de aparición de los síntomas y el tipo de síntoma observado varían en función de las condiciones medioambientales, del estado fisiológico de la planta en el momento en que se produce la inoculación por *B. tabaci*, así como del cultivar que se infecta. En infecciones realizadas en condiciones controladas (25°C, 16 h luz) los síntomas suelen aparecer en torno a los 15 días después de la inoculación. En tomate, en general, los síntomas incluyen reducción en la superficie foliary las hojas presentan el margen de los folíolos curvado hacia el haz a lo largo del nervio principal, lo que le da un aspecto de cuchara. Cuando las infecciones se producen en fases tempranas del crecimiento, la reducción del tamaño de los folíolos combinada con la parada del crecimiento en altura de la planta da a ésta un aspecto arbustivo. Las infecciones pueden ocasionar aborto floral y la abscisión de flores, falta de cuajado de frutos y producción de frutos no comerciales de pequeño tamaño y con poco desarrollo de color, llegando a pérdidas del 100% de la producción en infecciones tempranas. Una vez dentro de la célula vegetal, el virión es transportado al núcleo gracias a la cápside. Una vez allí se libera el ssDNA, que comienza a replicarse. La replicación de este ssDNA circular ocurre en tres fases: iniciación, elongación y terminación. La proteína Rep cataliza la iniciación y la terminación de la replicación por círculo rodante (RCR) a través del corte y la ligación de los extremos libres del DNA viral en un sitio de secuencia conservada en el *stem loop* de la zona intergénica (Laufs et al., 1995b). Tal como ocurre con el resto de virus ssDNA, los geminivirus no poseen ningún ORF que codifique para una polimerasa ni para elementos del replisoma en la horquilla de replicación (Hanley-Bowdoin et al., 2004; 2013). En plantas sanas, la disponibilidad de varios componentes

del replisoma está regulada por el ciclo celular por lo que debe ser reprogramado por el geminivirus para poder replicar su genoma.

Se ha demostrado que Rep es la única proteína esencial para la replicación de los geminivirus, por lo que se piensa que puede jugar un papel fundamental en el reclutamiento y ensamblaje del replisoma viral, incluyendo proteínas virales y proteínas del hospedador implicados en la replicación de DNA, reparación y otras funciones nucleares. REn interacciona con Rep incrementando notablemente la replicación y acumulación del DNA viral (Settlage et al., 2005). Ambas Rep y REn se unen a PCNA (Bagewadi et al., 2004; Castillo et al., 2003). Rep también interacciona con la subunidad

Replicación mediante círculo rodante (RCR)



Replicación dependiente de recombinación (RDR)

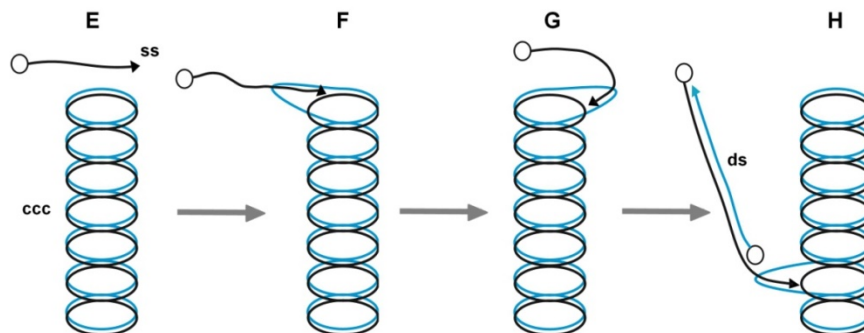


Figura 7. Modelos para la replicación por círculo rodante (RCR) y la replicación dependiente de recombinación (RDR) en begomovirus. Paso A: unión de la Rep al origen de replicación (ori) de un DNA de doble cadena circular abierto [(oc)dsDNA]. Paso B: corte del DNA y unión covalente del extremo 5' a la Rep. Paso C: replicación y desplazamiento del ssDNA. Paso D: corte de la nueva cadena y cierre de la molécula de ssDNA tras la separación de la Rep. Paso E: DNA de cadena sencilla (ssDNA) truncado interactúa en un punto homólogo con el DNA de cadena doble circular [(ccc)dsDNA] superenrollado. Paso F: recombinación homóloga. Paso G: migración del bucle y elongación del ssDNA. Paso H: elongación del ssDNA y síntesis de la cadena complementaria resultando en un DNA de cadena doble (dsDNA). Adaptada de: Jeske *et al.* (2001).

grande del factor C de replicación (RFC1) encargado de unir PCNA al DNA así como con la subunidad de 32 KDa de la proteína de replicación A, cuya función es la unión de ssDNA (Luque et al., 2002; Singh et al., 2007).

También se ha descrito la interacción entre Rep y RAD54, responsable de la recombinación homóloga por lo que podría jugar un papel importante en la replicación dependiente de recombinación (RDR). La interacción con RAD54 y con PCNA tiene efectos opuestos en la actividad de Rep *in vitro* por lo que ésta podría estar modulando RCR y RDR (Kaliappan et al., 2012; Luque et al., 2002). La **Figura 9** representa de forma esquemática los dos mecanismos aceptados para la replicación en geminivirus.

Una vez se ha replicado el genoma de los geminivirus, el dsDNA forma minicromosomas con 11 o 12 nucleosomas (Pilartz and Jeske, 1992). Rep también interacciona con la histona H3 (Kong and Hanley-Bowdoin, 2002), previniendo la metilación de la lisina 9 de la histona H3, implicada en la inhibición de la replicación del genoma viral (Viral Genome Methylation as an Epigenetic Defense against Geminiviruses, 2008).

5.4. TYLCV como modelo de estudio de cuasiespecies virales de ssDNA *in vivo*

El sistema experimental elegido para evaluar la importancia de las polimerasas TLS en la generación de variabilidad de virus ssDNA *in vivo* ha sido el del geminivirus TYLCV en *Arabidopsis thaliana*. Se trata de un virus emergente de ssDNA perteneciente a un grupo de virus con gran importancia agronómica y que pueden infectar una gran variedad de cultivos tanto de monocotiledóneas como de dicotiledóneas provocando cuantiosos daños económicos.

Se ha visto que estos virus evolucionan lentamente en ausencia de interacciones entre ellos (Sánchez-Campos et al., 2002) pero tienen un alta capacidad de generar nuevas variantes a través de procesos de recombinación cuando hay infecciones mixtas (García-Andrés et al., 2006; Monci et al., 2002b; Morilla et al., 2004; Navas-Castillo et al., 2000). Como se ha mencionado en la introducción general, en contra de lo que inicialmente se pensaba por tratarse de virus de DNA, los geminivirus presentan una alta variabilidad genética tal como ha demostrado para el CLCuV (Sanz et al., 2000) así como una estructura genética en sus poblaciones de tipo cuasispecies en los mastrevirus MSV (Isnard et al., 1998) y en el begomovirus TYLCCNV (Ge et al., 2007). Este hallazgo implica que la frecuencia de mutaciones puntuales en los geminivirus tiene una mayor

ocurrencia de la que se pensaba respecto al de la recombinación, y que muy probablemente esta estructura genética poblacional en cuasispecies sea la principal causa de la emergencia de nuevas especies de geminivirus. Recientemente se ha comprobado que alguno de los aislados de la especie TYLCV es capaz de infectar *A. thaliana* (Cañizares et al., 2015). Este resultado brinda la posibilidad de utilizar las herramientas genéticas y moleculares que se han desarrollado para esta planta modelo en estudios con este tipo de virus.

Materiales y Métodos

1. Material biológico

1.1. Plantas

Para la realización de este trabajo se han utilizado las estirpes y líneas mutantes comerciales de *A. thaliana* que aparecen en la tabla 1. Todas las líneas mutantes fueron obtenidas de la colección del *European Arabidopsis Stock Centre* a través de NASC (<http://arabidopsis.info/>).

Tabla 1. Líneas de *A. thaliana*

Estirpe/línea	Ecotipo	Observaciones
WT Col0	Columbia	Estirpe silvestre (<i>Wild type</i> , WT), utilizada como control por ser el ecotipo de <i>A. thaliana</i> del que derivan el resto de líneas mutantes empleadas en este trabajo.
Salk_011334C	Columbia	Línea mutante del gen At5g44750 (<i>atrev1-1</i>) por inserción de T-DNA en homocigosis. (descrita en (Jesús Santiago et al., 2008; Nakagawa et al., 2011b; Takahashi et al., 2005))
Salk_005721C	Columbia	Línea mutante del gen At5g44750 (<i>atrev1-2</i>) por inserción de T-DNA en homocigosis, (descrita en (Jesús Santiago et al., 2008; Richter et al., 2016c; Takahashi et al., 2005))
Salk_090163C	Columbia	Línea mutante del gen At5g52800 por inserción de T-DNA en homocigosis, ortólogo del gen humano <i>CCDC111</i> Primpol ((García-Gómez et al., 2013)) y que en este estudio se denominará <i>atprimpol</i>
Salk_129731C	Columbia	Línea mutante del gen At5g44740 (<i>atpolh</i>) por inserción de T-DNA en heterocigosis, (descrita en (Anderson et al., 2008; Jesús Santiago et al., 2008; Nakagawa et al., 2011b; Richter et al., 2016c)), por lo que se trata de una línea segregante para dicha inserción.
GK-485G11.(1-12)	Columbia	Línea mutante del gen At1g49980 por inserción de T-DNA, ortólogo del gen humano <i>POLK</i> . Tiene una segunda inserción de T-DNA en At2g42890. La línea consta de 12 lotes de semillas T ₃ segregantes para ambas inserciones. En este trabajo At1g49980 se denominará <i>atpolk</i>

1.2. Cepas bacterianas

Las cepas bacterianas empleadas para el desarrollo experimental de este capítulo pueden encontrarse en la tabla 2.

Tabla 2. Cepas bacterianas

Especie	Cepa	Genotipo	Observaciones
<i>Escherichia coli</i>	DH5 α	<i>fhuA2 lac(Δ)U169 phoA glnV44 Φ80' lacZ(Δ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17⁽¹⁾</i>	Cepa receptora utilizada para la clonación de fragmentos de DNA, en la propagación y el almacenamiento de plásmidos.
<i>Agrobacterium tumefaciens</i>	LBA4404	<i>TiAch5 pAL4404</i>	Cepa empleada para transformar plantas de <i>A. thaliana</i> . Contiene un plásmido coadyuvante con resistencia a rifampicina y estreptomicina

(1)http://ecoliwiki.net/colipedia/index.php/DH5_alpha

1.3. Clon infectivo

La agroinfección de *A. thaliana* se ha llevado a cabo con el clon infectivo de la cepa Mild del begomovirus TYLCV (TYLCV-Mld, p1.7SP72/97) (Navas-Castillo et al., 1999) amablemente cedido por el Dr. Enrique Moriones del Instituto de Hortofruticultura Subtropical y Mediterránea La Mayora (CSIC). Este clon es una construcción que consta de 1,7 copias de TYLCV-Mld (número de acceso del GenBank AF071228) en el plásmido binario pBin19 (**Figura 1**). De esta forma, cuando *Agrobacterium tumefaciens* transfiere al genoma de la planta el inserto de 1,7-mer se transcribe la proteína REP la cual escinde la secuencia de TYLCV situada entre las dos horquillas. Así se libera el monómero de TYLCV que tras su circularización puede comenzar la infección.

La cepa Mld puede infectar *A. thaliana* de forma asintomática. Dado que la infección es moderada, resulta un buen sistema para estudiar la infección a tiempos largos (Cañizares et al., 2015).

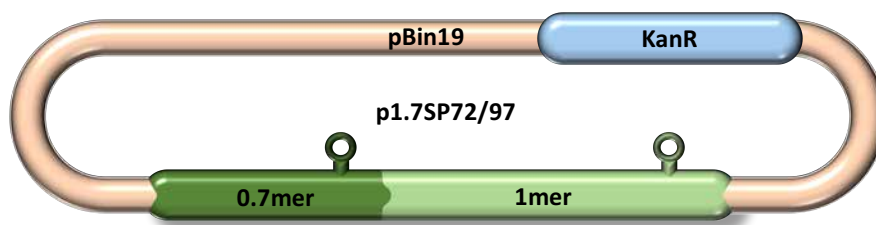


Figura 1. Representación esquemática del clon infectivo p1.7SP72/9. El clon contiene insertado una copia 1.7mer del genoma de TYLCV-Mld (segmento verde) en el plásmido binario pBin19 (naranja), que porta el gen de resistencia a kanamicina (KanR). *A. tumefaciens* integra el inserto al azar en el genoma de *A. thaliana*. El inserto de 1.7 mer incluye dos regiones IR con estructura de *stem-loop* con los que interacciona la proteína vírica Rep. Como consecuencia, se libera y promueve la circularización del monómero de TYLCV permitiendo la infección.

2. Medios de cultivo

2.1. Medios de cultivo bacteriano

Los medios de cultivo se esterizaron en autoclave (120°C, 20 min) o mediante filtración. La adición de antibióticos se realizó después del autoclavado y cuando el cultivo alcanzó temperaturas inferiores a los 60°C, en condiciones de esterilidad.

- Medio Luria-Bertani (LB)

Se usó para el crecimiento de *E. coli* o como base de otros medios de cultivo.

Líquido: bacto triptona (10 g/l), extracto de levadura (5 g/l) y NaCl (5 g/l).

Sólido: se añadió bactoagar (15 g/l) a la mezcla anterior previamente disuelta.

Para seleccionar y crecer *E. coli* transformada con p1.7SP72/97 se añadió al medio kanamicina (50 µg/ml)

- Medio para cultivo de *Agrobacterium tumefaciens*

Líquido: LB líquido suplementado con rifampicina (50 µg/ml) y estreptomicina (50 µg/ml).

Sólido: se añadió bactoagar (15 g/l) a la mezcla anterior previamente disuelta.

Para crecer *A. tumefaciens* transformado con p1.7SP72/97 se suplementa el cultivo con kanamicina (50 µg/ml).

- Medio PSI-b

Se usó para la preparación de células competentes.

Líquido: bacto triptona (20 g/l), extracto levadura (5 g/l), NaCl (0.6 g/l), KCl (0.185 g/l); ajustar el pH a 7. Después de esterilizar en autoclave se añadió 10 mM MgSO₄.

2.2. Medios para cultivo de *Arabidopsis thaliana*

Medios estériles

Los medios de cultivo se esterizaron en autoclave (120°C, 20 min) o mediante filtración. La adición de antibióticos se realizó después de la esterilización por autoclave y cuando el cultivo alcanzó temperaturas inferiores a los 60°C, en condiciones de esterilidad.

Medio MS+sacarosa

Medio nutritivo para la siembra de semillas en placas de Petri. Compuesto por: Murashige and Skoog Basal Medium (MS) (4.4 g/l), sacarosa (30 g/l), agar de plantas (8 g/l); Se ajustó el pH a 5.8 (con KOH).

Para la selección por fenotipo de líneas Salk se añadió a la mezcla previa kanamicina (50 µg/ml) después de haber sido esterilizada por autoclave.

Para la selección por fenotipo de líneas GABI-Kat se añadió a la mezcla previa sulfadiazina (50 µg/ml) después de haber sido esterilizada por autoclave.

Medios no estériles

Para el crecimiento de *A. thaliana* en condiciones no estériles se probaron dos medios de cultivo diferentes.

En un principio se usó una mezcla de turba, fibra de coco y litonita (6:3:1) y se crecieron en semilleros de 5 cm de lado. Posteriormente se optó por el sustrato comercial inerte prehidratado Jiffy® VECO1 de 20 semilleros por bandeja con el objetivo de trabajar en un sustrato lo más homogéneo posible.

3. Cultivo y crecimiento de organismos

3.1. Cultivo y crecimiento de bacterias

Escherichia coli se cultivó en placas de Petri con medio LB sólido, o en matraces y/o tubos de ensayo con una capacidad 4 veces mayor que el volumen de cultivo para permitir una buena aireación del mismo, con medios de cultivo apropiados y en agitación (175-200 rpm). La temperatura de incubación fue de 37°C.

A. tumefaciens se creció también en medio LB, suplementado con antibióticos cuando procediese, a 28°C con agitación (200 rpm) o en medio sólido en estufas, a la misma temperatura.

La conservación de las estirpes bacterianas durante largos periodos de tiempo se llevó a cabo a -80°C en medio LB líquido con 15% glicerol (v/v). Para periodos cortos, se mantuvieron a 4°C en medio LB sólido.

3.2. Cultivo y crecimiento de plantas

- Cultivo de plantas en macetas

La germinación de *A. thaliana* se llevó a cabo en semilleros con turba previamente humedecida en la que se sembraron un número aproximado de 100 semillas por línea. A continuación los semilleros se cubrieron con plástico transparente para favorecer el aumento de humedad en el semillero.

Para sincronizar la germinación de las semillas, los semilleros se incubaron a 4°C durante 48 horas en oscuridad. Tras este tiempo, se colocaron en cámaras de crecimiento con tubos fluorescentes de 36 W, 8 horas de luz, 16 horas de oscuridad, es decir, en ciclo de día corto para favorecer el desarrollo vegetativo y a una temperatura de 22 ±1°C.

Transcurrida una semana (o tras la aparición de los cotiledones), se retiró el plástico y se mantuvieron en las mismas condiciones hasta que alcanzaron la fase de 4 hojas verdaderas de más de 2 mm, momento en el cual se trasplantaron a macetas individuales.

En un primer experimento para el crecimiento de las líneas mutantes de *A. thaliana* se empleó como sustrato una mezcla de turba, fibra de coco y litonita (6:3:1), comúnmente utilizada en el laboratorio, pero esta composición produjo crecimientos aberrantes de la parte aérea de la planta, cuyas causas desconocemos (**Figura 2**). Por ello finalmente se decidió optar por el sustrato comercial Jiffy® VECO1 que resultó óptimo para el crecimiento de estas líneas mutantes tal como se muestra en la **figura 3**. En este sustrato las plantas se regaron mediante subirrigación con agua destilada cada dos días.

2





Figura 3: Fenotipo de las líneas WT, *atrev1-2*, *atpolh*, *atrev1-1*, *atpolk* y *atprimpol* con 30 días, crecidas en Jiffy® VECO1. Todas las líneas mostraron fenotipo silvestre.

En el caso en que las plantas fueran destinadas a producir semillas para favorecer el desarrollo floral se empleó un ciclo de día largo (16 horas de luz y 8 horas de oscuridad) en cámaras de crecimiento con tubos fluorescentes de 36W a 24°C. Cuando las plantas dejaron de producir flores se suspendió la irrigación y se dejaron secar en bolsas de

papel. Las semillas se recogieron y se tamizaron con coladores convencionales. Su conservación se realizó a temperatura ambiente, en tubos eppendorf o falcon cerrados, en cajas de plástico rodeados de gel de sílice.

- **Cultivo en placas de Petri**

Previo a la siembra en placa, las semillas se esterilizaron sumergiéndolas primero en una solución de lejía al 50% (v/v) y de otra de 0,1% Tween 20 (v/v), manteniéndolas en agitación durante 15 min. Posteriormente, se realizaron 4 lavados consecutivos de 5 min cada uno con agua destilada estéril. Por último se le añadió una solución de agarosa al 0,2% estéril para facilitar la siembra individual en placa. Para la siembra se tomaron de forma sucesiva pequeñas fracciones alícuotas de la solución acuosa que contenía las semillas esterilizadas con una pipeta y se dispensaron en placas de Petri con medio MS y sacarosa al 3%. Las placas se sellaron con papel de parafina y se incubaron en una cámara con tubos fluorescentes (36 W, 8 horas de luz, 16 horas de oscuridad) a una temperatura de $22\pm 1^{\circ}\text{C}$ durante 15 días de forma que estas pudiesen desarrollar su fenotipo resistente o sensible al antibiótico.

4. Oligonucleótidos

Los oligonucleótidos cebadores (Sigma Aldrich) utilizados como cebadores en reacciones de amplificación y/o cuantificación, se diseñaron con la ayuda de los programas Primer3 y OligoCalc, analizando en cada caso su estabilidad interna, la formación de dímeros y horquillas no deseados, así como diversos parámetros físicos y químicos (T_m , T_d , %G+C, %A+T).

4.1. Cebadores para el genotipado de las líneas mutantes de *A. thaliana*

Las secuencias de todos los cebadores empleados para el genotipado de las líneas mutantes se obtuvieron de la web del *Salk Institute* (signal.salk.edu/tdnaprimers.2.html) tal como se muestra en la **Tabla 3**.

Para caracterizar el genotipo de las líneas mutantes fue necesario realizar dos reacciones de PCR. En la primera reacción se usó la pareja de cebadores específica de cada línea mutante denominados LP, si se trataba del cebador *forward*, o RP, si se trataba del *reverse*. Estos cebadores flanquean la región del DNA de *A. thaliana* donde se ubica la inserción de T-DNA. En el caso de que no exista un inserto de T-DNA, el amplificado de la PCR con esta pareja de cebadores tendrá un tamaño aproximado de 1 Kb, indicando

que al menos un alelo del gen analizado es silvestre, tal como se muestra en la **Figura 4A** para las líneas Salk y la **Figura 4C** para las líneas Gabi-Kat. En este capítulo, para diferenciar entre homólogos de estas polimerasas utilizaremos el prefijo “at” para referirnos a loci, mutantes o proteínas pertenecientes a *A. thaliana* y el prefijo “h” para referirnos a los humanos.

Tabla 3. cebadores empleados para el genotipado de las líneas mutantes de *A. thaliana*

Línea	Nombre	Gen	Secuencia
atrev1-1 (Salk_011334C)	atrev1-1 LP atrev1-1 RP	At5g44750	5'-ACCTTTTCTCCAGAGGTGCTC 5'-GAGTCCTCCTCACATGCAAG
atrev1-2 (Salk_005721C)	atrev1-2 LP atrev1-2 RP	At5g44750	5'-CTCTTCAAAAGGATTTTGGGG 5'-CGTTCAGTCTCAAGGACCAAC
atrimpol (Salk_090163C)	atrimpol LP atrimpol RP	At5g52800	5'-TTCACCTTCGTCACTTTGATGAG 5'-ATTCAACATTTTCTTGCTCCG
atpolh (Salk_129731C)	atpolh LP atpolh RP	At5g44740	5'-GAGGACTGCCTAGTGCAAGTTG 5'-GTCAGTTTGCAAGCTAGTGCC
atpolk (GK-485G11)	atpolk LP atpolk RP mei2 LP mei2 LP	At1g49980 At2g42890	5'-AACGATGTGCAATGGAGAGAC 5'-TATGGGCAACCTCATTCTTTG 5'-CATGAAAAAGGAGGGCCTAAC 5'-AAGAATGCGTGTATTGGGATG
Línea T-DNA	Nombre	Secuencia	
Salk	LBb 1.3	5'-ATTTTGCCGATTTTCGGAAC	
Gabi-Kat	o8474	5'- ATAATAACGCTGCGGACATCTACATTTT	

La segunda reacción de PCR difiere entre líneas Salk y Gabi-Kat. Para las líneas Salk se utilizó la pareja de cebadores compuesta por el cebador RP (específico de cada línea mutante) y el LBb1.3 que reconoce el extremo 3' del T-DNA. Si existiese un T-DNA inserto el producto de amplificación tendría un tamaño de entre 500 y 800 nt (**Figura 4B**). Para la línea Gabi-Kat se utilizó la pareja de cebadores compuesta por el oligonucleótido LP (específico de cada línea mutante) y el cebador o8474, que reconoce el extremo 5' del T-DNA. El producto de amplificación resultante, al igual que en las líneas Salk, tendría un tamaño de entre 500 y 800 pb si existiese un T-DNA inserto como se indica en la **Figura 4D**. Esta reacción de PCR indicará si existe al menos un alelo en el gen diana con una inserción de T-DNA, que junto con los resultados de la primera PCR, indicará si la línea es homocigota o heterocigota.

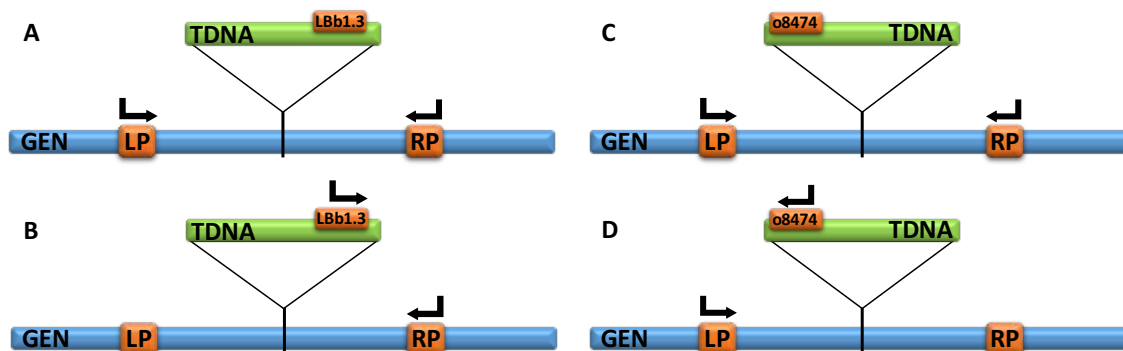


Figura 4: Parejas de cebadores utilizados para el genotipado de las líneas mutantes de *A. thaliana*. En azul se muestra el gen objeto posible diana de la inserción de TDNA (verde). En naranja están señaladas las secuencias reconocidas por los diferentes cebadores utilizados en el genotipado. En el caso de las líneas Salk una amplificación positiva utilizando cebadores LP y RP (A) indica que al menos uno de los alelos tiene genotipo silvestre, mientras que una amplificación positiva con la pareja de cebadores Lb1.3 y RP (B) indica la presencia de TDNA en al menos uno de los alelos del gen. En el caso de las líneas Gabi-Kat una amplificación positiva utilizando cebadores LP y RP (C) indica que al menos uno de los alelos tiene genotipo silvestre. Una amplificación positiva con la pareja de cebadores o8474 y LP (D) indicará la presencia de TDNA en al menos uno de los alelos del gen.

4.2. Cebadores para la cuantificación relativa de mRNA de *Arabidopsis thaliana*

Todos los cebadores utilizados para el estudio de expresión relativa fueron diseñados para cumplir los estándares de calidad necesarios para qPCR, a saber, T_m entre 59,8°C y 60,2°C; %GC entre el 40 y el 60%, producto de amplificación entre 100 y 200 bp; longitud de cada oligo entre 18 y 21 nt y no más de 3 G/C en 3'OH ni en 5'OH. Todos los cebadores usados en la RT-qPCR han presentado una eficiencia (E) entre el 98% y el 102% y una R^2 superior a 0,995. La eficiencia es un valor que relaciona la pendiente de la curva patrón obtenida en la reacción de comprobación con su óptimo teórico a través de la fórmula $E=10^{-1/m}$, donde m el valor de la pendiente. Para la cuantificación relativa de la expresión se aplicó el método $2^{-\Delta\Delta C_T}$ que no tiene en cuenta la eficiencia de la reacción y estima siempre que es del 100%, por ello nos aseguramos de que la eficiencia de la reacción se encuentre en valores muy cercanos al 100%. En cuanto al valor R^2 hace referencia al ajuste de los puntos medidos en una reacción de qPCR con respecto a la curva patrón, de forma que valores cercanos a 1 indican un ajuste óptimo, e inferiores hacen referencia a variabilidad entre réplicas.

Para la reacción de retrotranscripción previa al estudio de la expresión relativa de los transcritos de los genes sometidos a estudio se empleó como cebador oligo dT.

En el caso del análisis de la expresión del gen *atrev1* se comprobó si existían diferencias de expresión entre ambos mutantes del mismo gen debido a las diferencias de posición del inserto de T-DNA. Para ello se diseñaron dos parejas de cebadores (**Tabla 4**), una pareja de cebadores aguas arriba y otra pareja aguas abajo de ambas inserciones de T-DNA de las líneas *atrev1-1* y *atrev1-2*.

Tabla 4. Cebadores empleados para la cuantificación relativa de DNA Polimerasas TLS en *Arabidopsis thaliana*

Gen	Cebador	Secuencia
ATREV1	qREV1-DOWN F	5'-CCTGATGCAAGTGCCAGTG
	qREV1-DOWN R	5'-CCACATTGTCCTGCAAGTATA
	qREV1-UP F	5'-GCAGTGCAAGTGCTGGAATA
	qREV1-UP R	5'-ACCAGGTAGTGTCCTGACTG
ATPRIMPOL	qPRIMPOL F	5'-ACGCTGCTGGTACTGGTTCT
	qPRIMPOL R	5'-TTCCCCTTCTGAGATCAACG
ATPOLH	qPOLH F	5'-AAAAGAATCGGGGAATGCCTTC
	qPOLH R	5'-CCGGTGTTAACTGCTTATTGG
ATPOLK	qPOLK F	5'-AGCCAAAGAAGCAGAAACCA
	qPOLK R	5'-GGAATTGGACGGTGAGAGAA
ACTINA2	AtACTIN F	5'-AAAGGATGCTTATGTTGGCG
	AtACTIN R	5'-GAAAGAGTAACACGCTCGG
	Oligo dT	5'-TTTTTTTTTTTTTTTTTTTT

4.3. Cebadores para la cuantificación absoluta de la carga viral de TYLCV

Para cuantificar la carga viral de las plantas de *A. thaliana* infectadas se utilizó la pareja de cebadores qTYLCV 835 F y qTYLCV 836 R (**Tabla 5**), que amplifican una región de 171 pb de la CP. Como normalizador de la cuantificación absoluta se utilizaron cebadores para el gen RNAr 25S de *A. thaliana*.

Tabla 5: Cebadores empleados para la cuantificación por qPCR absoluta de TYLCV Mild

Organismo	Cebador	Secuencia
TYLCV	qTYLCV 835 F	5'-GAAGGCTGAACCTCGACAGC
	qTYLCV 836 R	5'-GGACTTTACATGGGCCTTCAC
<i>A.thaliana</i>	25S_Universal-F	5'-ATAACCGCATCAGGTCTCCA
	25S_Universal-R	5'-CCGAAGTTACGGATCCATTT

5. Extracción y purificación de ácidos nucleicos

5.1. Purificación de DNA plasmídico

La purificación de plásmidos se realizó siguiendo las instrucciones de Illustra plasmidPrep Mini Spin Kit® (GE Healthcare Life Sciences) siguiendo las instrucciones del fabricante. El procedimiento está basado en una lisis alcalina seguida de la unión del DNA plasmídico a una matriz de sílice y su posterior elución con 50 µl de buffer Tris-HCl 10 mM a pH 8.

5.2. Extracción de DNA de *Arabidopsis thaliana* infectadas con TYLCV

Para la extracción de DNA, todas las plantas de *A. thaliana* se pulverizaron en un mortero con Nitrógeno líquido y todo el material biológico fue almacenado en tubos eppendorf de 1.5 ml a -80°C hasta su uso.

Para cada muestra se partió de 150 mg de tejido pulverizado que se introdujo en un tubo eppendorf de 1.5 ml al que se añadieron 400 µl de tampón de extracción (200 mM Tris-HCl, pH 7,5, 200 mM NaCl, 25 mM EDTA, 0,5%, 0,1% β-mercaptoetanol). El tejido se trituró con ayuda de un pistilo de plástico y el contenido del tubo se centrifugó a 12000 g durante 5 min a temperatura ambiente. El sobrenadante se transfirió a un nuevo tubo al que se le añadieron 300 µl de fenol:cloroformo 1:1 (v/v) equilibrado a pH 8, mezclando bien las fases antes de usarlo. A continuación se centrifugó a 12000 g durante 5 min a temperatura ambiente y la fase acuosa fue depositada en un nuevo tubo. El DNA se precipitó con 300 µl de isopropanol tras incubación durante 2 min a temperatura ambiente. Tras centrifugar a 12000 g durante 5 min el precipitado se lavó una vez con etanol al 100% y otra con etanol al 70%. Por último, el DNA se secó a temperatura ambiente y se resuspendió en 100 µl de tampón TE (10 mM Tris HCl pH 8, 1 mM EDTA). En caso de no ser usado inmediatamente, el DNA fue almacenado a -20°C.

5.3. Extracción de RNA total de *Arabidopsis thaliana*

Se partió de 150 mg de muestra pulverizada en Nitrógeno líquido en un tubo de eppendorf de 1.5 ml. A cada muestra se le añadieron 300 µl de Tri Reagent Solution (Ambion), se agitó la muestra durante 2 min y se centrifugó a 12000 g durante 5 min a 4°C, pasando el sobrenadante a un nuevo tubo. A este nuevo tubo se le añadieron 60 µl de cloroformo, se agitó durante 15 s y se dejó reposar 3 min a temperatura ambiente. A continuación se centrifugó a 12000 g durante 15 min a 4°C. El sobrenadante se transfirió

a un tubo nuevo al que se añadieron 150 µl de isopropanol. Después de 10 min de reposo se volvió a centrifugar a 12000 g 10 min a 4°C. El sobrenadante se transfirió a un nuevo tubo al que se le añadieron 300 µl de etanol al 70% y se mezcló mediante vortex 15 s. Las muestras se centrifugaron 6000 g durante 5 min a 4°C, se descartó el sobrenadante y se dejó secar el pellet a temperatura ambiente. Finalmente el pellet se resuspendió en un volumen de 25 µl de agua MilliQ previamente tratada con dietilpicrocarbonato 1:10000 (v:v) (agua DEPC).

6. Tratamiento enzimático de ácidos nucleicos

6.1. Tratamiento con endonucleasas de restricción

Las enzimas de restricción empleadas fueron suministradas por la casa comercial TaKaRa. Los tratamientos del DNA con una única enzima de restricción o con enzimas que actúan en el mismo tampón se llevaron a cabo a la temperatura recomendada durante al menos 1 h utilizando 1-2 unidades de enzima por microgramo de DNA. Cuando los tratamientos se realizaron con dos enzimas que no actúan en el mismo tampón se seleccionó el tampón óptimo recomendado por la casa para digestiones dobles de endonucleasas de restricción. Cuando la restricción se realizó con dos enzimas cuyas temperaturas óptimas de reacción eran diferentes, en primer lugar, se realizó la digestión con la enzima cuya temperatura óptima era más alta para, a continuación, incubar con la de temperatura más baja. Para cada temperatura el tiempo de incubación fue de al menos 1 h. Las enzimas de restricción se inactivaron según las instrucciones del fabricante, generalmente incubándolas durante 30 min a 65°C.

- Tratamiento de las extracciones de DNA total con DpnI

Todas las muestras de DNA procedentes de plantas agroinfectadas pasaron por un tratamiento con la endonucleasa *DpnI*. Esta enzima de restricción reconoce un sitio GATC sólo si las adeninas están metiladas por Dam ($G_{m6}ATC$) una metilasa presente en *A. tumefaciens*. De esta forma, con el tratamiento de *DpnI* se digirió el DNA del plásmido conteniendo secuencias de TYLCV procedentes de *A. tumefaciens* y no el DNA viral circular que se haya replicado en el interior de la planta infectada.

6.2. Eliminación de endonucleasas de las muestras

Para la purificación del DNA tratado con endonucleasas se utilizó el kit comercial StrataClean Resin (Agilent Technologies), siguiendo las instrucciones indicadas por el

fabricante. El funcionamiento de este kit se basa en una purificación libre de fenoles en la cual, la fase sólida de resina contiene grupos hidroxilos que reaccionan con las proteínas con la misma afinidad que lo harían los grupos hidroxilos del fenol, permitiendo una separación del DNA y las endonucleasas. Para ello, en todos los casos se añadió un volumen de resina del doble de volumen de la muestra y se mezcló con vórtex durante 15 segundos. La mezcla se incubó 1 min a temperatura ambiente y se centrifugó a 2000 g durante otro minuto. El sobrenadante, conteniendo el DNA, fue transferido a un nuevo tubo y se repitió el proceso desde el principio una vez más, obteniendo el mismo volumen inicial de la muestra.

7. Reacciones de amplificación de ácidos nucleicos

7.1. PCR estándar

Para la amplificación estándar de fragmentos de DNA se empleó el kit BIOTAQ™ DNA polimerasa termoestables de BIOLINE. La reacción contenía 1,5 pmol/μl de cada cebador, 4 mM de MgCl₂, 0,2 mM de dNTP y 1 U de polimerasa. La cantidad de DNA molde empleada fue de 200 ng. Las condiciones de reacción se indican en la **Tabla 6**.

Tabla 6. Etapas de la reacción en cadena de la polimerasa (PCR)

Etapa	Temperatura	Tiempo	Nº de ciclos
Desnaturalización inicial	94°C	5 min	1
Desnaturalización	94°C	30 s	N
Unión al cebador	55°-65°C ^a	30 s	
Extensión	72°C	T_{ext} ^b	
Extensión final	72°C	10 min	1
Mantenimiento	10°C	Indefinido	1

^aLa temperatura es dependiente de la T_m del cebador, por regla general se usó la relación T=T_m-4°C.

^bT_{ext} el tiempo de extensión de la reacción, directamente proporcional al tamaño amplificado, fue de 30 s por Kb amplificada. N fue de 30 ciclos excepto en las PCR para genotipado de plantas que fue de 40 ciclos.

7.2. PCR para genotipado de plantas

Para caracterizar el genotipo de las líneas mutantes de *A. thaliana* se emplearon las parejas de cebadores descritas en el apartado 4.2 de Materiales y Métodos. De esta forma, la PCR generó dos amplicones de distintos en tamaño: un fragmento que indica la presencia del alelo silvestre y/o uno correspondiente al alelo mutante causado por la inserción de DNA-T. Las condiciones de la reacción fueron las indicadas en el anterior.

Debido al número elevado de plantas que hubo que caracterizar de forma simultánea, se optó por el uso del kit KAPA3G Plant PCR (KAPA Biosystems) siguiendo las instrucciones del fabricante. Este método permitió llevar a cabo las reacciones de PCR de genotipado directamente a partir de extractos crudos de hojas sin necesidad de purificar previamente el DNA genómico, gracias a la alta procesividad y resistencia a inhibidores de la polimerasa incluida en el kit, la DNA polimerasa KAPA3G Plant.



Figura 5. Genotipado con el kit Harris®. El sacabocados permitió obtener trozos de tejido de 0,5 mm. Fotografía realizada en el laboratorio de Genética de la Facultad de Ciencias de la Universidad de Málaga.

Los extractos de hoja se obtuvieron usando el kit para extractos de hojas Harris® (Harris Inc.) con un sacabocados con un diámetro de 0.5 mm tal como se muestra en la figura 5. Para prevenir la contaminación cruzada entre muestras, entre cada una de las tomas de muestras se esterilizó el extremo del sacabocados con lejía comercial y posteriormente se lavó con agua-DEPC.

7.3. Síntesis de DNAc mediante transcripción inversa

Se utilizó el kit *AMV Reverse Transcriptase* (Promega) junto con el kit *Recombinant RNasin Ribonuclease Inhibitor* (Promega) y se siguieron las indicaciones del fabricante. A una dilución que contenía 2 µg de RNA total se le añadieron 0,25 µl de cebador a 10 µM en un volumen final de 11 µl. A continuación las muestras se calentaron 5 min a 70°C y se mantuvieron 5 min en hielo. Posteriormente se le añadió una mezcla que contenía 5 µl de tampón de reacción, 2,5 µl de dNTPs (10 mM cada uno), 40 U de *RNasin Ribonuclease inhibitor* y 30 U de AMV RT. La reacción consistió en un ciclo de 60 min a 42°C y un ciclo de 10°C de forma indefinida.

7.4. PCR cuantitativa

Antes de llevar a cabo las reacciones de qPCR, se cuantificó cada muestra de DNA tanto por electroforesis en gel de agarosa como por espectrofotometría. Para las reacciones de qPCR siempre se utilizaron 5 réplicas biológicas y 3 réplicas técnicas para cada una de ellas. La reacción se llevó a cabo en un termociclador Bio-Rad iCycler 3 en un volumen final de 20 µl, conteniendo 2 µl de DNA o DNAC diluido a 130 ng/µl, 16 µl de SsoFast™ EvaGreen® Supermix (Bio-Rad) y 1 µl de cada uno de los cebadores. Las condiciones de la PCR se detallan en la **Tabla 7**.

Tabla 7. Cebadores empleados para la cuantificación por qPCR absoluta de TYLCV Mild

Organismo	Cebador	Secuencia
TYLCV	qTYLCV 835 F	5'-GAAGGCTGAACTTCGACAGC
	qTYLCV 836 R	5'-GGACTTTACATGGGCCTTCAC
<i>A. thaliana</i>	25S_Universal-F	5'-ATAACCGCATCAGGTCTCCA
	25S_Universal-R	5'-CCGAAGTTACGGATCCATT

- Cuantificación relativa

Esta técnica se empleó para estudiar el nivel relativo de transcrito de los genes *ATREV1*, *ATPOLH*, *ATPOLK* Y *ATPRIMPOL* en plantas silvestres y las líneas mutantes *atrev1-1*, *atrev1-2*, *atpolh*, *atpolk* y *atprimpol* en relación a las silvestres. Los cebadores utilizados están descritos en el apartado 4.3.

La normalización de los datos se llevó a cabo basándose en los valores de cuantificación de ciclo (C_q , *quantification cycle*) resultantes de la amplificación por triplicado de la expresión del gen de referencia *ACTIN2*. Este normalizador fue utilizado para llevar a cabo el método $2^{-\Delta\Delta C_T}$ (Bustin et al., 2009).

- Cuantificación absoluta

Para la cuantificación absoluta de TYLCV-Mild, se prepararon curvas patrón usando el plásmido pBluescript II SK (+) conteniendo una copia completa del genoma de TYLCV-Mild, donado amablemente por el Dr. Eduardo Rodríguez Bejarano de la Universidad de Málaga. Con el fin de hacer una recta patrón, se realizaron diluciones decimales seriadas desde 10^9 a 10^3 copias por µl del plásmido. Para el cálculo del número de moléculas de doble cadena de DNA a partir de la concentración de DNA de cada muestra se empleó la fórmula:

$$n^{\circ} \text{ de copias de dsDNA} = \frac{N \cdot 6,022 \times 10^{23}}{T \cdot 1 \times 10^9 \cdot 650}$$

donde N=ng/μl y T=pares de base del fragmento de DNA.

Para cada pareja de cebadores se obtuvieron curvas estándar por análisis de regresión lineal de los valores de C_q de tres réplicas de las diluciones estándar sobre el logaritmo de la cantidad de DNA total de cada muestra.

La normalización de los datos de cuantificación absoluta para cada muestra se llevó a cabo basándose en los valores C_q resultantes de la amplificación por triplicado del gen de referencia correspondiente al codificante para el RNA ribosómico 25S con los cebadores 25S_Universal-F y 25S_Universal-R siguiendo el protocolo descrito por Mason et al. (2008). Las condiciones de la reacción y tiempos están reflejados en la **tabla 8**.

Tabla 8. Condiciones de PCR cuantitativa absoluta (qPCR).

Etapa ^a	Temperatura	Tiempo	N° de ciclos
Desnaturalización inicial	95°C	30 s	1
Desnaturalización	95°C	10 s	40
Unión al cebador/Extensión	60°C ^b	15 s	
Curva de disociación	60°C+0.5°C/ciclo	10 s	60

^aLa recogida de datos de fluorescencia se efectúa en las etapas de unión al cebador y en cada ciclo de la curva de disociación (*melting*).

^bTodas las parejas de cebadores fueron diseñados para usar la misma temperatura de unión (*annealing*).

7.5. Amplificación por círculo rodante (RCA)

Para amplificar de forma específica el genoma viral presente en los extractos de DNA total, se utilizó la amplificación por círculo rodante (*Rolling circle amplification*, RCA), con el kit IllustraTM ThempliPhi (GE Healthcare). Este kit utiliza la DNA polimerasa del bacteriófago Φ29, cuya característica principal es su elevada procesividad y alta fidelidad de copia, para amplificar de forma preferente genomas circulares a partir de cebadores con secuencias aleatorias de 6 nucleótidos. Para ello, se siguieron las indicaciones del fabricante. Inicialmente en un tubo de polipropileno de 0,5 ml con 5 μl de *Sample Buffer* se le añadió 1 μl de DNA extraído de las plantas infectadas a una concentración de 130 ng/μl y se calentó a 95°C durante 3 min. Las muestras se enfriaron durante 5 min en hielo y se les añadieron 5 μl de una mezcla de 5 μl de *Reaction Buffer*

y 0,2 µl de *Enzyme mix* por cada muestra. Se mezcló la reacción con cuidado y se centrifugó 15 s a 1800 g para concentrar toda la muestra en el fondo del tubo. A continuación las reacciones se mantuvieron a 30°C durante 8 horas y finalmente la reacción de RCA fue inactivada por calentamiento de las muestras a 65°C durante 10 min.

Para la secuenciación de las muestras de RCA, se realizaron dos amplificaciones para cada una de las muestras y se unió el producto de ambas amplificaciones en un volumen final de 20 µl. El hecho de utilizar dos reacciones independientes de RCA para una sola muestra se justifica por la necesidad de reducir la acumulación de errores aleatorios que pueda producir Φ29 en cada una de las reacciones, ya que los errores generados en una reacción no serán idénticos a los que se producen en la segunda, consiguiendo de esta forma que el ratio mutaciones reales/mutaciones RCA se incremente al doble.

8. Electroforesis de ácidos nucleicos

8.1. Electroforesis en geles de agarosa

La separación e identificación de moléculas de DNA se realizó mediante electroforesis horizontal en geles de agarosa (Agarose SPI, Duchefa) disuelta en tampón TBE (Tris-Base 89 mM, EDTA 2 mM y Ácido Bórico 89 mM). La concentración de agarosa varió en función del tamaño de los fragmentos a separar, oscilando entre 0,8-1% (p/v). A la solución de agarosa fundida se añadió Sybr Safe (1 µl/ml) para la posterior visualización de los ácidos nucleicos. Cada muestra se mezcló con tampón de carga para geles de agarosa en proporción 5:1 (v/v). La electroforesis se realizó en cubetas horizontales con tampón TBE, a un voltaje constante de 3-4 V/cm. Los geles se fotografiaron con un documentador de geles ChemiDoc XRS+ (Biorad) empleando el programa Image Lab®. Como patrones de tamaño molecular se empleó 1 Kb Plus DNA Ladder de Invitrogen. La concentración de DNA se determinó midiendo la densidad óptica (DO) a una longitud de onda de 260 nm con un espectrofotómetro NanoVue Plus (GE Healthcare) o por cuantificación en geles de agarosa en comparación con patrones de DNA de cantidades conocidas.

9. Transformación genética

9.1. Preparación de células competentes

- Células competentes por choque térmico de *Escherichia coli*

Los cultivos se inocularon a partir de una colonia fresca de *E. coli* DH5 α en 5 ml de PSI-b (ver apartado 2.1. de Materiales y Métodos) y se incubó toda la noche con agitación a 37°C. Todo el volumen del cultivo en fase estacionaria se transfirió a 100 ml de PSI-b en un matraz de 1 L (precalentado a 37°C) y se continuó la incubación con agitación vigorosa hasta alcanzar una DO₆₀₀ de 0,45-0,5. A continuación, el cultivo se enfrió en hielo y se centrifugó a 5.000 g durante 5 min a 4°C. Seguidamente, las células se resuspendieron suavemente en 40 ml de solución Tfb-1 [30 mM acetato potásico pH 6,9, 50 mM MnCl₂, 100 mM KCl, 100 mM CaCl₂, 15 % glicerol (v/v)] fría. Tras una incubación de 5 min en hielo, las células se volvieron a centrifugar a 5.000 g durante 5 min a 4°C, se retiró el sobrenadante y se resuspendieron suavemente en 4 ml de solución Tfb-2 [10 mM MOPS pH 7, 75 mM CaCl₂, 10 mM KCl, 15% glicerol (v/v)] fría. Por último, esta suspensión celular se incubó en hielo durante 15 min y se distribuyó en fracciones alícuotas de 150 μ l que se congelaron rápidamente en N₂ líquido y se almacenaron a -80°C hasta su uso.

- Células competentes por electroporación de *Agrobacterium tumefaciens*

Los cultivos se inocularon a partir de una colonia fresca de *A. tumefaciens* LBA4404 siguiendo las indicaciones de (An, 1987) en 10 ml de LB suplementado con los antibióticos rifampicina (50 μ g/ml) y estreptomycin (50 μ g/ml) e incubaron durante la noche a 28°C con agitación a 175 rpm. A continuación, se transfirieron 100 μ l del pre-inóculo a 500 ml de LB líquido y se incubó 18 horas en agitación, hasta que se alcanzó una DO₆₀₀ de 0,88. Finalmente se centrifugó a 6000 rpm durante 10 min, se descartó el sobrenadante y el pellet de bacterias se resuspendió en 250 μ l de agua MilliQ estéril. Este último paso de centrifugación y resuspensión del pellet se repitió 3 veces más para finalmente resuspenderlo en glicerol al 10% (v/v). El volumen final de bacterias *A. tumefaciens* competentes se dividió en fracciones alícuotas de 100 μ l que se conservaron durante un tiempo máximo de 6 meses a -80°C.

9.2. Transformación de *Escherichia coli*

A 50 μ l células competentes se añadieron 5 μ l del clon infectivo p1.7SP72/97 y se incubaron 30 min en hielo. Transcurrido ese tiempo, las células se sometieron a un choque térmico de 2 min a 42°C e inmediatamente se volvieron a dejar en hielo durante

5 min. Por último, las células se incubaron a 37°C con agitación durante 30 min en 1 ml de LB. Tras centrifugar, las células se resuspendieron en 150 µl de medio LB. Por último toda la mezcla de transformación se sembró en placas de medio selectivo que se incubaron a 37°C.

9.3. Transformación de *Agrobacterium tumefaciens*

50 µl de *A. tumefaciens* competentes se transfirieron a una cubeta de electroporación junto con 5 µl del clon infectivo purificado de *E. coli* tal como se indica en el apartado 5.1. La cubeta con la muestra de bacterias y el plásmido se introdujeron en un electroporador Eppendorf Modelo 2510 y se le dio un pulso de 2500 V a la muestra. Las células transformadas fueron transferidas a un tubo eppendorf de 1,5 ml con 1 ml de LB y se incubaron a 28°C durante 2 horas. Tras centrifugar, las células se resuspendieron en 150 µl de medio LB. Por último toda la mezcla de transformación se sembró en placas de medio selectivo que se incubaron a 28°C.

9.4. Agroinfección de *A. thaliana*

La agroinfección se llevó a cabo mediante la técnica de punteadura del ápice (Cañizares et al., 2015). Para ello se utilizaron siempre plantas de *A. thaliana* 30 días postrasplante, cuando la roseta alcanzaba más de 3 cm de diámetro. A partir de un cultivo en césped de *A. tumefaciens* transformado con el clon infectivo p1.7SP72/97 se cogió biomasa con la punta de un palillo de madera estéril y se depositó en el ápice de la planta y a continuación se picó el ápice 10 veces con una aguja entomológica. Finalmente se limpió el ápice de los restos de *A. tumefaciens* con una botella de lavado con agua destilada para evitar que la bacteria pudiese producir necrosis apical.

10. Secuenciación

La secuenciación de las muestras infectadas amplificadas por RCA fue llevada a cabo por la empresa MacroGen, en Seul (Corea del Sur). Se realizó una primera secuenciación con pocas muestras con la tecnología Illumina HiSeq que produce secuencias de hasta 100 nt. Las genotecas se realizaron con el kit Truseq DNA PCR-free, con un tamaño de inserto de 330 pb. Las muestras de DNA amplificado por RCA se prepararon a una concentración de 20 ng/µl en un volumen final de 100 µl. El número de lecturas recibido por MacroGen ascendió a 4.5×10^7 secuencias pareadas en cada una de las 5 muestras. El resto de las muestras secuenciadas por NGS se secuenciaron mediante la tecnología Illumina MiSeq.

Para ello las genotecas se generaron con el kit Truseq DNA PCR-free con un tamaño de inserto de 550 nt. Las muestras de DNA se suministraron a Macrogen a una concentración de 100 ng/μl en un volumen final de 20 μL. La tecnología de Illumina MiSeq permitió obtener lecturas pareadas de una longitud de 301 nt cada una. Las 23 muestras se secuenciaron en una sola carrera, obteniendo finalmente un número de lecturas pareadas en bruto de entre $1,0 \times 10^6$ y $1,8 \times 10^6$ en cada muestra.

11. Bioinformática

11.1. Uso del programa QuasiFlow

Las lecturas de secuencia proporcionadas por Macrogen fueron analizadas con el programa de diseño propio QuasiFlow, explicado en detalle en el Capítulo 2 de esta Tesis, utilizando las variables descritas en la **Tabla 9**. QuasiFlow caracteriza muestras de poblaciones víricas, lo que permite analizar la estructura molecular de la población de virus en un hospedador concreto y se basa en las siguientes etapas: limpieza y separación de lecturas, análisis estructural de las secuencias, reconstrucción de haplotipos y finalmente recopilación/graficado de los resultados. El resultado generado por Quasiflow proporciona información sobre las recombinaciones, SNP, indel, haplotipos, transiciones, transversiones y calcula el índice de Shannon normalizado (Sn) y la diversidad nucleotídica (*d*).

Tabla 9. Variables empleadas en el análisis de secuencias de TYLCV con QuasiFlow

Variable	Valor
<i>Secuencia de referencia</i>	TYLCV-Mild (<i>GenBank accession</i> AF071228)
<i>Posición de comienzo de la secuencia consenso</i>	1
<i>Posición final de la secuencia consenso</i>	2791
<i>Secuencia circular o lineal</i>	C
<i>Probabilidad de aceptación de recombinaciones</i>	0.000007
<i>ORFs^a</i>	155l116ldir; 315l258ldir; 164l359lrev; 321l100lrev; 1152l135lrev; 1300l134lrev
<i>Posiciones nucleotídicas de interés 1:</i>	-
<i>Posiciones nucleotídicas de interés 2:</i>	-
<i>Valor del tamaño del chunk^b</i>	5000 (MiSeq) 20000 (HiSeq)

^aLos valores en esta variable siguen la estructura x|y|z; donde "x" es en nt de inicio del ORF, "y" es el número de codones que contiene el ORF, "z" es la dirección desde la que enumerar los nt: "dir" en sentido horario y "rev" en sentido antihorario.

11.2. Programas para la gestión de datos y análisis de secuencias

Para el procesado y manejo de los datos obtenidos en el transcurso del trabajo así como para el análisis de las secuencias estudiadas, se han utilizado distintos programas informáticos (Tabla 10).

Tabla 10. Programas empleados para la gestión y análisis de datos y secuencias

Programa	Empresa	Aplicación
QuasiFlow	Universidad Málaga	Tratamiento y análisis comparativo de secuencias
CFX Manager 3.0	Bio-Rad	Análisis de datos de PCR cuantitativa
SigmaPlot		Análisis estadísticos.
Microsoft Excel 15.16	Microsoft	Cálculos y análisis estadísticos.
APE v2.0.49		Edición de secuencias, análisis de sitios de restricción
Tablet 1.15.09.01	Hutton Institute	Estudio de alineamientos de NGS
ImageLab 4.1	Bio-Rad	Procesado y edición de imágenes de geles de agarosa
UCSF Chimera 1.10.01	Univ. of California	Modelado y análisis de estructuras 3D de proteínas
Gephi 0.8.2	NetBeans	Diseño de redes de interacción
Geneious 9	Biomatters	Filogenia, edición de secuencias, sitios de restricción
Primer3		Diseño de oligonucleótidos
OligoCalc		Comprobación de cebadores para PCR cuantitativa

Resultados

1. Caracterización de las líneas mutantes de *A. thaliana atrev1-1*, *atrev1-2*, *atpolh*, *atpolk* y *atprimpol*

Hemos planteado la hipótesis de que al menos parte de la variabilidad que caracteriza a las cuasiespecies de virus de ssDNA pueda deberse a la acción de las DNA polimerasas de translesión (TLS). Las cadenas sencillas de DNA, como ocurre en el caso de TYLCV, serían más susceptibles a sufrir ciertos daños como desaminación, formación de 8-oxo-dG o dímeros de timina que las cadenas dobles de DNA, y, por lo tanto, podrían ser sustrato de las DNA Polimerasas de TLS y responsables de la generación de la variabilidad observada en estos virus.

Para analizar la variabilidad de poblaciones virales de TYLCV en ausencia de estas DNA polimerasas de TLS empleamos las líneas mutantes de *A. thaliana atrev1-1*, *atrev1-2*, *atpolh*, *atpolk* y *atprimpol*. Estas líneas fueron adquiridas comercialmente de la web del *European Arabidopsis Stock Centre* (<http://arabidopsis.info>). Las líneas mutantes de *A. thaliana atrev1-1*, *atrev1-2*, *atpolh* y *atprimpol* pertenecen a la línea comercial Salk, creadas en el *Salk Institute*(CITA) , mientras que la línea mutante de *A. thaliana atpolk* pertenece a la línea comercial Gabi-Kat (GK), desarrolladas en la Universität Bielefeld (CITA). Ambas líneas (Salk y GK) son mutantes por inserción de T-DNA en los loci correspondientes a estas TLS (*ATREV1-1*, *ATREV1-2*, *ATPOLH* Y *ATPOLK*) y a *PRIMPOL*, tan sólo se diferencian en el T-DNA que llevan inserto cada una de estas líneas. En el caso de las líneas Salk, el T-DNA incluye un gen de resistencia a kanamicina, mientras que la línea GK incluye el gen de resistencia a la sulfadiacina. El hecho de tener T-DNA diferentes hace que también su genotipado se efectúe con diferentes cebadores que reconozcan los diferentes T-DNA. La línea Salk recomienda el uso de un cebador diseñado por el *Salk Institute* llamado LBb 1.3, mientras que para la línea GK, sus autores diseñaron hasta 17 cebadores distintos que reconocen el T-DNA, de los cuales se seleccionó el llamado o8474, porque funcionalmente es muy similar al cebador LBb 1.3 de la línea Salk.

En el caso de las líneas Salk (*atrev1-1*, *atrev1-2*, *atpolh* y *atprimpol*), se ha descrito la posibilidad de que el gen de resistencia al antibiótico que incluye el T-DNA haya sido silenciado (XIII *International Meeting on Arabidopsis Research*, Sevilla, 2002)

imposibilitando la selección por fenotipo de estas líneas mutantes. Además para las líneas *atrev1-1*, *atrev1-2*, *atpolh* y *atpolk*, no se ha comunicado la presencia o ausencia del gen de resistencia activo. Por ello, como trabajo previo, caracterizamos tanto fenotípicamente como genotípicamente todas las líneas de *A. thaliana* seleccionadas para este estudio.

1.1. Fenotipado de las líneas mutantes de *Arabidopsis thaliana*

Las líneas Salk *atrev1-1*, *atrev1-2*, *atpolh* y *atprimpol* y las 12 sublíneas Gabi-Kat *atpolk* se crecieron en medio MS suplementado con kanamicina y sulfadiazina respectivamente para comprobar si aún conservaban la expresión de los genes de resistencia que contenían las inserciones de T-DNA de estas líneas. Todas las líneas Salk presentaron sensibilidad al antibiótico y murieron una semana después de germinar, mostrando decoloración de los cotiledones e inhibición del crecimiento justo después de la germinación. Dicho fenotipo se muestra en la **Figura 1** para el caso de la línea *atrev1-1*.

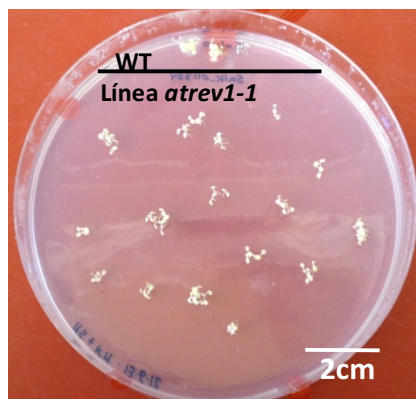


Figura 1. Germinación de semillas WT y de la línea *atrev1-1*. Se muestra una placa de MS sólido suplementado con kanamicina con semillas germinadas de plantas WT en la parte superior, y en el resto de la placa, las semillas de la línea *atrev1-1*. Estas presentaron una sensibilidad a la kanamicina similar al control. El resto de líneas Salk (*atrev1-2*, *atpolh* y *atprimpol*) presentaron la misma sensibilidad al antibiótico

La línea Gabi-Kat *atpolk* 485g11 se trata de una línea T_2 (el término T_2 hace referencia a una descendencia F_2 procedente de un F_0 a la que le fue insertada una o más copias de T-DNA por agroinfección) segregante para 2 inserciones de T-DNA en los genes *ATPRIMPOL* y *MEI2*. Por ello se recibieron 12 sublíneas T_3 numeradas del 1 al 12 procedentes de dicha T_2 . Las sublíneas Gabi-Kat 485g11.01, 485g11.02, 485g11.03, 485g11.04, 485g11.05, 485g11.06, 485g11.07, 485g11.08, 485g11.09, 485g11.10,

485g11.11 y 485g11.12 segregantes T₃ de Gabi-kat 485g11 (*atpolk*) sí mostraron resistencia a la sulfadiazina. Por simple herencia mendeliana existe la posibilidad de relacionar el porcentaje de semillas resistentes de cada línea con el número de loci afectados por inserciones de T-DNA. Por ello en primer lugar comprobamos en el caso de las sublíneas *atpolk* si había en cada una de ellas 1 locus o dos loci afectados, ya que estas presentan dos inserciones de T-DNA segregantes, una de ellas en *ATPOLK* y la segunda en *MEI2*. Puesto que todas las líneas GK-485g11 (1-12) resultaron tener un porcentaje de germinación entre el 50% y el 82% en presencia de sulfadiazina, ninguna de las líneas pudo ser descartada como portadora de una o dos inserciones en el genoma (Tabla 1).

Tabla 1: Resistencia a sulfadiazina de las líneas *atpolk*

GK-485g11 (<i>atpolk</i>)	Semillas germinadas	Semillas Resistentes	% de semillas resistentes	Nº de loci probables con inserción de T-DNA
1	33	27	82%	2
2	40	20	50%	1 ó 2
3	30	15	50%	1 ó 2
4	23	13	57%	1 ó 2
5	148	107	72%	2
6	72	36	50%	1 ó 2
7	47	27	57%	1 ó 2
8	77	59	77%	2
9	81	67	83%	2
10	164	120	73%	2
11	218	142	65%	2
12	101	70	69%	2

Dado que en ninguno de los casos analizados el fenotipo se pudo emplear para establecer el genotipo de las diferentes líneas mutantes, se procedió a su genotipado y al análisis de expresión.

1.2. Genotipado de las líneas mutantes de *Arabidopsis thaliana*

El genotipado fue realizado tal como se indicó en el apartado 4.1. de Materiales y Métodos del presente capítulo. Para las líneas Salk *atprimpol*, *atrev1-1*, *atrev1-2* y *atpolh* se genotiparon 40 plantas en dos reacciones independientes, una con la pareja de cebadores LP y RP específica para cada línea y otra con la pareja de cebadores RP y LbB1.3 (Figura 2).

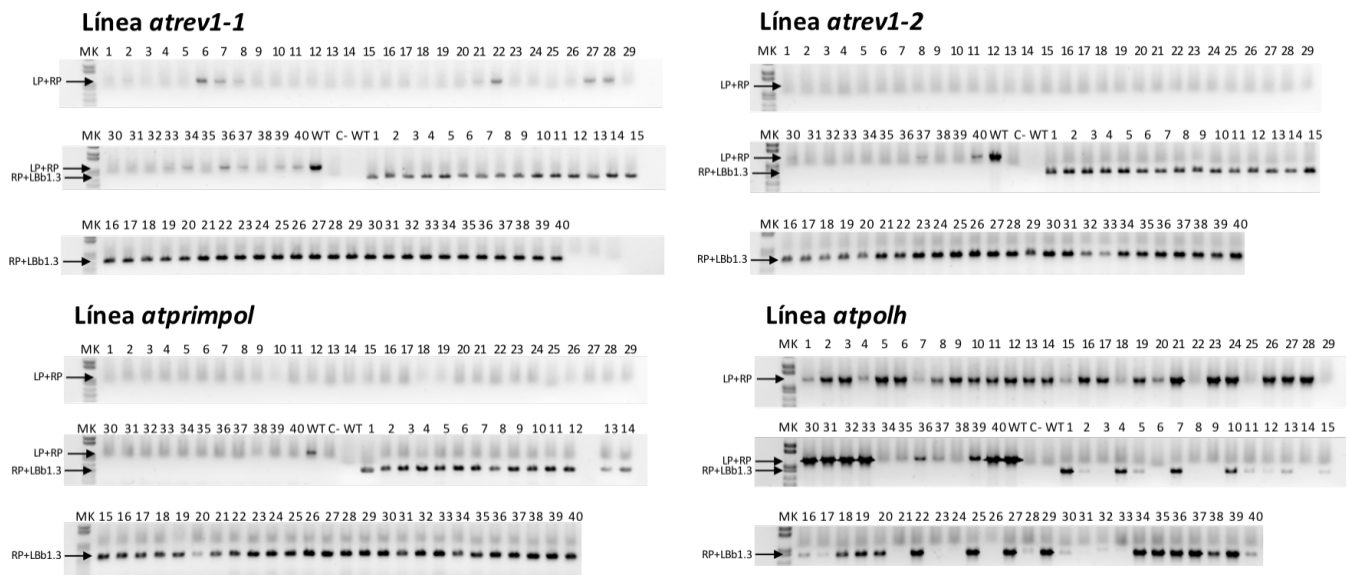


Figura 2. Genotipado de las líneas mutantes *atrev1-1*, *atrev1-2*, *atprimpol* y *atpolh*. Electroferogramas de geles de agarosa al 1% en los que se muestran los productos de la amplificación de dos reacciones de PCR separadas: la primera con los cebadores específicos RP+LP para cada de las líneas, cuya amplificación da un producto en torno a las 1000pb, muestra el genotipo silvestre; y la segunda con la pareja de cebadores RP y LBb1.3, que da lugar a un amplificado de entre 550 y 850 pb (dependiendo de la línea), revela el genotipo mutante por inserción de T-DNA. Para cada línea mutante se analizaron 40 plantas. Las líneas *atrev1-1*, *atrev1-2* y *atprimpol* resultaron ser homocigotas para la inserción de TDNA mientras que la línea *atpolh* resultó ser heterocigota segregante

En el caso de las líneas *atprimpol*, *atrev1-1* y *atrev1-2*, la pareja de cebadores LP y RP sólo amplificaron en el control positivo, excepto en las plantas 37 y 40 de *atrev1-1* y las plantas 6, 7, 8, 22, 27, 28, 34, 36, 39 y 40 de la línea *atrev1-2*. Para estas plantas dudosas se realizaron dos nuevas PCR aumentando la temperatura de unión del cebador al DNA a 55°C y a 60°C y así comprobar si la amplificación de la PCR se debía a una unión inespecífica de los cebadores. Tal como se observa en la figura 3, el producto de amplificación de unos 1000 nt que se observa en todas las plantas cuando la hibridación de los cebadores se realizaba a 50°C desapareció al incrementar dicha temperatura a 60°C. Dado que se mantenía la amplificación en el control se dedujo que se trataban de una amplificación inespecífica. La segunda PCR de genotipado de las líneas *atprimpol*, *atrev1-1* y *atrev1-2* realizada con la pareja de cebadores RP y LBb1.3 fue positiva en todas las réplicas analizadas por PCR, lo que indicó que eran homocigotas para la inserción de T-DNA.

Por el contrario, la línea *atpolh* resultó ser heterocigota para la inserción de T-DNA (**Figura 2**), ya que la pareja de cebadores LP y RP dio lugar a un producto de amplificación en más del 50% de las muestras analizadas, al igual que con la pareja de cebadores RP y LBb1.3.

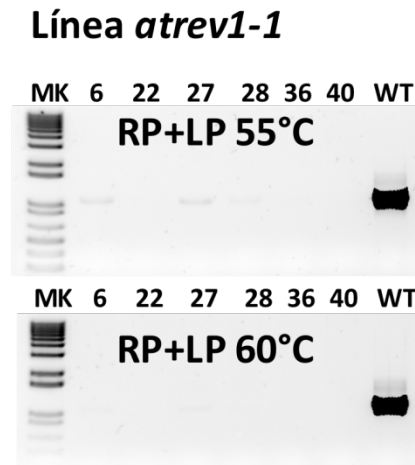


Figura 3. Comprobación de muestras de genotipo dudoso en la línea *atrev1-1*. Electroferogramas de dos geles de agarosa al 1% en el que se muestran los productos de la amplificación de las muestras 6, 22, 27, 28, 36 y 40 de la línea *atrev1.1* que mostraron una posible amplificación con los cebadores específicos RP y LP a 50°C. En el electroferograma superior se observa un producto de amplificación de 1000 pb que desaparece al incrementar la temperatura de *annealing* de los cebadores de 50° a 60°C, tal como se ve en el electroferograma de abajo..

En el caso de las 12 líneas T₃ segregantes con dos inserciones de TDNA, Gabi-Kat-485g11, se optó por el uso de los cebadores *atpolk*_RP, *atpolk*_LP y o8474 en una sola reacción para analizar la inserción de T-DNA en *PolK* y los cebadores *mei2*_LP, *mei2*_RP y o8474 para para genotipar la inserción en *MEI2* de estas líneas y se comprobó que la línea T₃ GK-485g11.06 era homocigota para *atpolk* y heterocigota para la inserción en *MEI2* tal como se muestra en la **Figura 4**.

Para la obtención de una línea que solo tuviera la inserción en *ATPOLK* se generaron semillas T₄ a partir de la línea GK-485G11.06. Una vez crecidas varias plantas de esta T₄, fueron genotipadas con los cebadores *mei2*_LP y *mei2*_RP (850 nt aproximadamente) en busca del fenotipo silvestre para *MEI2*. La **Figura 5** muestra uno de los geles de agarosa donde se pudo identificar alguna de estas plantas silvestres para *MEI2*. Estas plantas se dejaron crecer hasta obtener semillas T5 GK-485G11.06 *atpolk* / *MEI2*

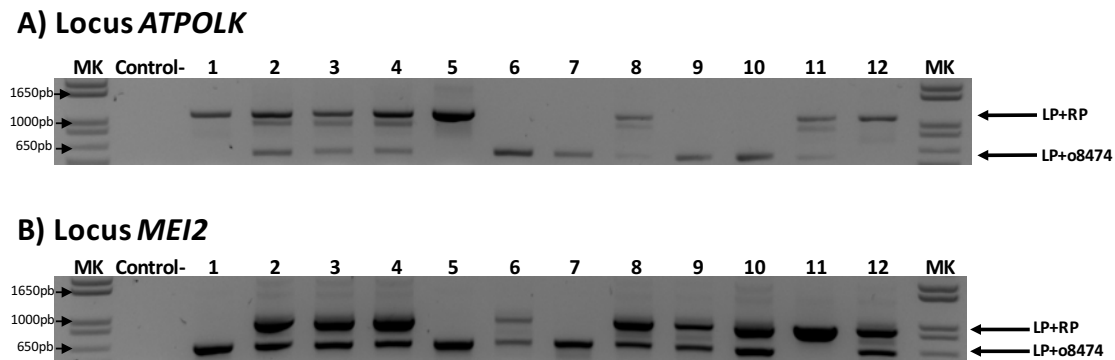


Figura 4. Genotipado de las líneas Gabi-Kat (1-12). Electroferograma de un gel de agarosa al 1% en el que se comprobó la existencia de dos inserciones de T-DNA en 2 regiones del genoma de *A. thaliana*. A) Genotipado del locus *ATPOLK*, con el objetivo de buscar homocigotos para la inserción de T-DNA. Las líneas 6, 7, 9 y 10 son homocigotas para la inserción ya que no hubo producto de amplificación con la pareja de cebadores LP+RP en dichas muestras. B) La única línea silvestre para el locus *MEI2* fue la línea 11 que era también silvestre para el locus *ATPOLK*. Las líneas 2, 3, 4, 6, 8, 9, 10 y 12 mostraron genotipo heterocigoto para *MEI2*. Con el fin de obtener una descendencia silvestre para *MEI2* se obtuvieron semillas de la línea 6, homocigota para *ATPOLK* y segregante para *MEI2*.

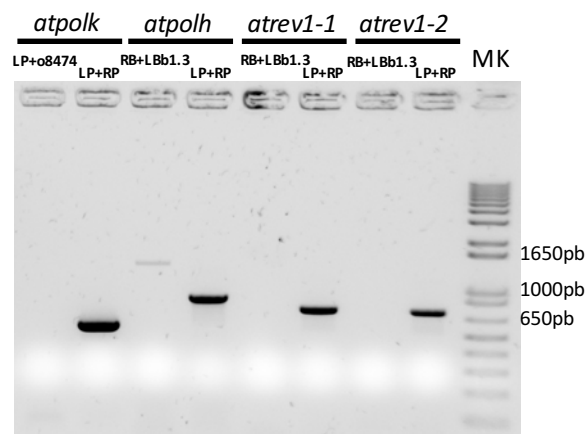


Figura 5. Genotipado de las líneas utilizadas para el estudio de la carga viral de TYLCV y secuenciación masiva. Electroferograma de un gel de agarosa al 1% utilizado para comprobar el genotipo de las líneas de *A. thaliana* mutantes para las polimerasas de TLS AtPolK, AtPolH, AtRev1.1 y AtRev1.2 que se van a infectar con TYLCV. La línea *atpolh* resultó ser portadora de la inserción en heterocigosidad (calles 3 y 4), mientras que el resto de líneas son homocigotas para la mutación de sus respectivos genes.

1.3. Expresión relativa de las diferentes DNA polimerasas TLS sobre fondos genéticos mutantes de *Arabidopsis thaliana*

Para comprobar el efecto de las inserciones de T-DNA en cada una de las líneas mutantes *atpolh*, *atpolk*, *atrev1-1*, *atrev1-2* y *atprimpol*, analizamos los valores de expresión relativa de los genes *ATREV1*, *ATPOLH* Y *ATPOLK* en todas las cepas en estudio y así

determinar si en ausencia de expresión de una de DNA Pol TLS en estudio alteraba la expresión del resto. En el caso de las líneas *atrev1-1* y *atrev1-2*, dado que contienen el inserto de T-DNA en posiciones diferentes del gen, se analizó adicionalmente la expresión de *ATREV1* en la región genómica inmediatamente anterior al extremo 5' del inserto de cada T-DNA en ambas líneas, para comprobar si los niveles de expresión diferían entre ambas líneas. Cada medida de la expresión relativa se obtuvo a partir de 4 réplicas biológicas con 3 réplicas técnicas cada una de ellas. Todos los datos se normalizaron a los niveles de expresión del gen endógeno *ACTIN2* y se relativizaron a la expresión en el *WT*. Todas las medidas fueron analizadas mediante t de Student para $P < 0,05$.

En el caso de la línea mutante *atpolh* (**Tabla 2**) se observó que los niveles de expresión del gen *ATPOLH* fue similar a los niveles de la expresión en el *WT* ($1,1 \pm 0,13$) no mostrando diferencias significativas entre ambos para un T de Student ($P < 0,05$). Este resultado sugiere que una sola copia del gen *ATPOLH* es suficiente para mantener los niveles de expresión tan altos como en el control. El análisis de la expresión de los genes *ATPOLK* y *ATREV1* de la línea heterocigota *atpolh* mostró que no había diferencias significativas con respecto al control *WT*. El descenso en la expresión de *ATPOLK* ($0,49 \pm 0,09$) no es estadísticamente significativo según la T de Student para $P < 0,05$. En el caso de la línea *atpolk*, los valores de expresión de *ATPOLK* se sitúan en $0,41 \pm 0,06$ veces menor que el gen silvestre, lo que indicó un descenso significativo de la expresión del mismo ($p < 0,05$). Aunque los valores de expresión no fueron lo suficientemente bajos para considerarla una línea *knockout*, se consideró una línea *knockdown* para *ATPOLK*. La expresión de *ATPOLH* y de *ATREV1* no se vio influida en esta línea, con valores de $0,87 \pm 0,09$ y $0,8 \pm 0,07$, respectivamente.

Las líneas *atrev1-1* y *atrev1-2* presentaron una ausencia total de expresión de *ATREV1* aguas abajo de las inserciones. Para *atrev1-1* el nivel de expresión fue de $0,01 \pm 0,005$ y en el caso de *atrev1-2* de $0,02 \pm 0,004$ con respecto al control ($p < 0,05$), por lo que ambas líneas mutantes se consideraron *knockout* para el gen *ATREV1*. No hubo cambios en los valores de expresión del resto de TLS en ninguna de las dos líneas mutantes.

Cuando se analizó la expresión del gen *ATREV1* en la región anterior al extremo 5' del T-DNA en las líneas *atrev1-1* y *atrev1-2*, se observó que en el caso de *atrev1-2* los niveles de transcrito descendían hasta $0,26 \pm 0,05$ ($p < 0,01$) mientras que en la línea *atrev1-1* los

valores de expresión se mantenían en niveles similares al del WT ($0,90 \pm 0,19$). Estas diferencias en los valores de expresión entre ambas líneas mutantes *atrev1-1* y *atrev1-2* indica la posibilidad de que la línea *atrev1-1* pudiera ser algo funcional, debido a que tiene el dominio catalítico intacto y se ha observado expresión de la misma.

Puesto que AtPrimPol no es una DNA Pol TLS no se consideró analizar la expresión de estas sobre el fondo mutante *atprimpol*. En cuanto al análisis de la expresión de *ATPRIMPOL* los valores fueron de $0,13 \pm 0,03$ con respecto al control ($p < 0,01$). Por lo que también se consideró a esta línea *knockout* para *ATPRIMPOL* tal como se observa en la **Tabla 2**.

Tabla 2. Expresión relativa de las diferentes polimerasas de translesión en los diferentes fondos genéticos mutantes de *A. thaliana*

Línea de <i>A. thaliana</i>	Expresión de ARNm relativo al WT				
	<i>ATPOLK</i>	<i>ATPOLH</i>	<i>ATREV1 3'</i>	<i>ATREV1 5'</i>	<i>ATPRIMPOL</i>
<i>atpolk</i> (homocigota)	0,41*±0,06	0,87±0,09	0,8±0,07	N.A.	N.A.
<i>atpolh</i> (heterocigota)	0,49±0,09	1,1±0,13	0,72±0,06	N.A.	0,92±0,03
<i>atrev1-1</i> (homocigota)	0,85±0,18	1,06±0,11	0,01**±0,005	0,90±0,19	1,75±0,27
<i>atrev1-2</i> (homocigota)	0,65±0,09	0,52±0,03	0,02**±0,004	0,26**±0,05	1,24±0,18
<i>atprimpol</i> (homocigota)	N.A.	N.A.	N.A.	N.A.	0,13**±0,03

* $p < 0,05$; ** $p < 0,001$; N.A.: no analizado.

Se resalta en gris el nivel de expresión del gen con el inserto de T-DNA

2. Cinética de la carga viral de TYLCV sobre fondos genéticos mutantes de DNA polimerasas TLS

Tal como se indicó en la introducción, las DNA polimerasas TLS no son polimerasas procesivas y reemplazan a las DNA polimerasas replicativas cuando existe algún tipo de daño en el DNA que pueda causar bloqueo de la replicación. De ser esta hipótesis correcta, sería de esperar un efecto en la carga viral como consecuencia de la existencia de genomas virales con daños en el DNA que no pueda ser reparado por una DNA Pol específica. Por ello decidimos analizar la carga viral de TYLCV en ausencia de PrimPol y de las DNA Pol TLS Rev1, PolH y PolK.

Para analizar la cantidad de moléculas virales que podía acumularse en cada fondo genético carente de alguna de las DNA Pols TLS estudiadas inoculamos el clon infectivo de TYLCV a un total de 20 plantas de 30 días de cada una de las líneas WT y los mutantes *atprimpol*, *atrev1-1*, *atrev1-2*, *atpolh*, *atpolk*. A los 7, 14, 21 y 28 días se recogieron 5 plantas de cada una de las líneas a las que se les extrajo el DNA total para estudiar la carga viral mediante qPCR absoluta normalizando los valores con respecto al gen *rRNA* 25S (Figura 6).

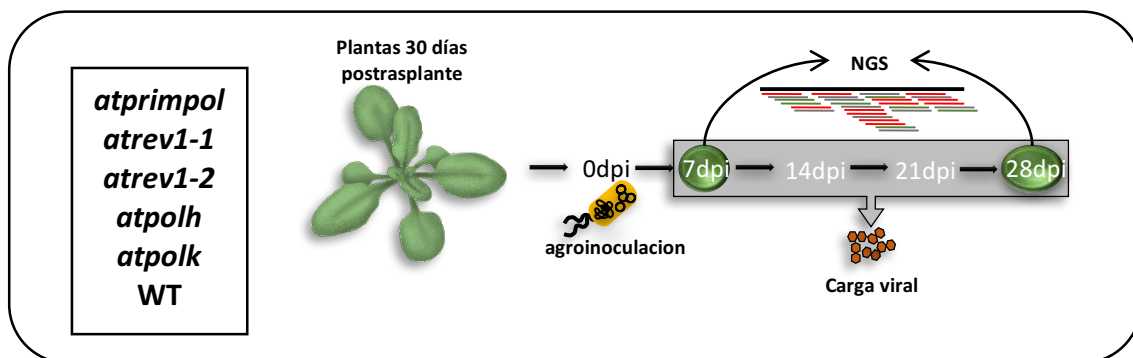
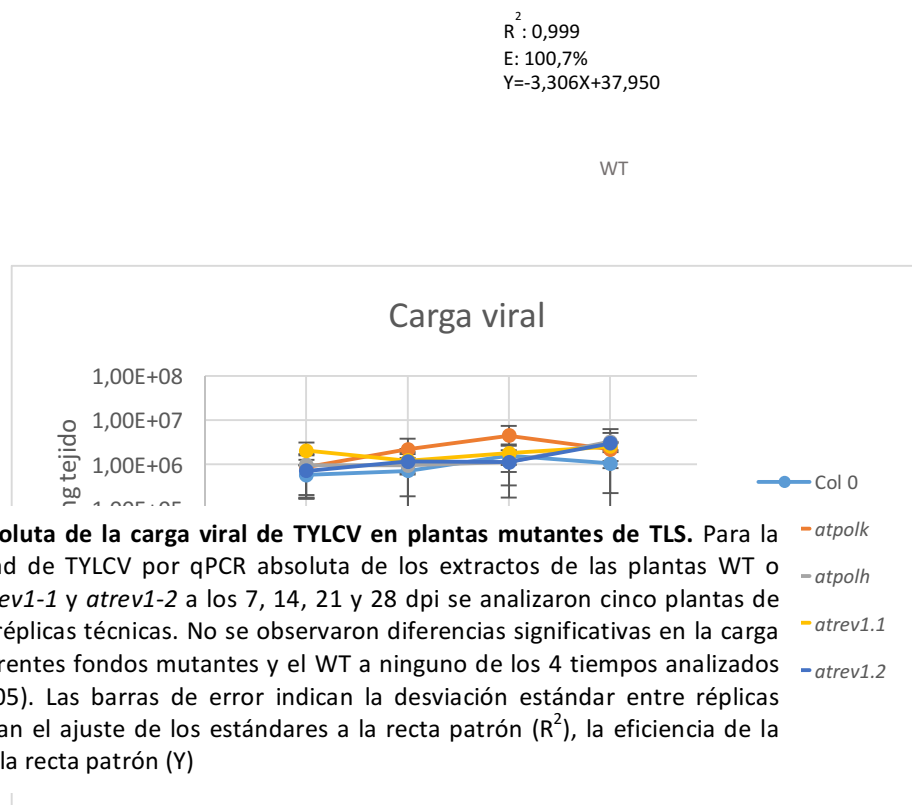


Figura 6. Esquema del diseño experimental para el estudio de la influencia de las TLS en la generación de variabilidad en poblaciones virales de ssDNA. A 20 plantas de 30 días de cada línea de *A. thaliana* (indicadas en el cuadro) se les inoculó el clon infectivo p1.7SP72/97 mediante *A. tumefaciens*. Tras 7, 14, 21 y 28 dpi se tomaron cuatro muestras a las que se analizó la carga viral. Se amplificaron por RCA todas las muestras y se analizaron los espectros de mutantes de cada línea infectada de 2 muestras a 7 dpi y 3 muestras a 28 dpi mediante secuenciación masiva HiSeq o MiSeq de Illumina.

La carga viral de TYLCV en las líneas mutantes de TLS (*atrev1-1*, *atrev1-2*, *atpolh* y *atpolk*) alcanzó valores máximos en el control y en todas las líneas mutantes a 7dpi (Figura 7). Estos valores oscilaron entre $5,69 \times 10^5$ y $2,10 \times 10^6$ moléculas de TYLCV por mg de tejido y se mantuvieron constantes a los 14, 21 y 28 días en todas las muestras. No se observaron diferencias significativas de niveles de expresión entre el WT y ninguna de las líneas mutantes *atrev1-1*, *atrev1-2*, *atpolh* o *atpolk* para $p < 0,05$.

En el caso de la línea mutante *atprimpol* se observó una disminución significativa de la carga viral a 7 dpi ($p < 0,01$), como muestra la figura 8. Esta diferencia osciló entre $1,47 \times 10^7$ moléculas de TYLCV en el control y $1,40 \times 10^6$ moléculas de TYLCV en la línea *atprimpol*. La desviación estándar entre réplicas biológicas osciló entre 0,88 en el caso del control y 0,22 en el caso de *atprimpol*. A los 14 y 21 dpi la carga viral se igualó en ambas infecciones, con valores entre $7,12 \times 10^6$ y $8,25 \times 10^6$ moléculas de TYLCV.



Los resultados sugieren que Rev1, PolH y PolH no están implicadas en la replicación de TYLCV en *A. thaliana* posiblemente debido a la baja procesividad que tienen estas DNA Pol TLS. Sin embargo la ausencia de PrimPol produjo un descenso del 90% de la carga viral a 7 dpi con respecto al WT, es decir, 9×10^6 moléculas de TYLCV menos, lo que sugiere su implicación como primasa en el inicio de la replicación viral como muestra la **Figura 8.**

$$R^2 : 0.996$$

$$E: 97.3\%$$

$$Y=-3.387X+37.124$$

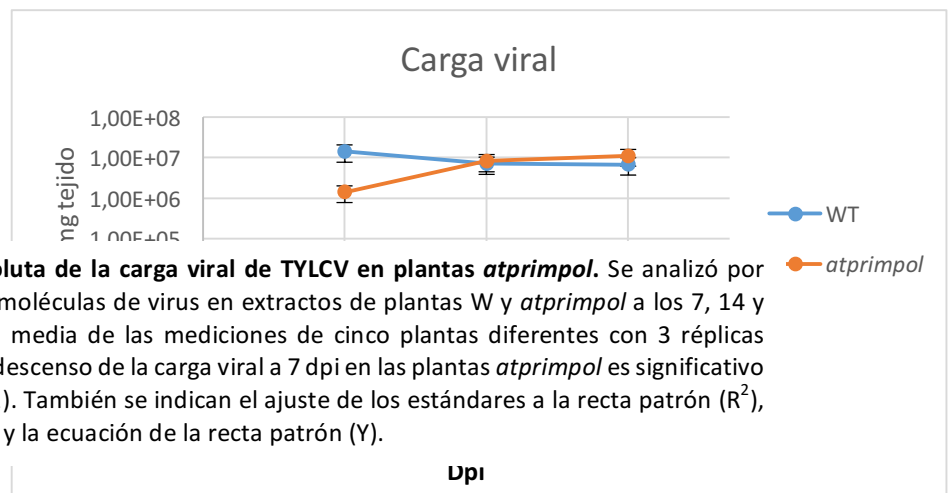


Figura 8. Cuantificación absoluta de la carga viral de TYLCV en plantas *atprimpol*. Se analizó por qPCR absoluta el número de moléculas de virus en extractos de plantas W y *atprimpol* a los 7, 14 y 21 dpi. Los resultados son la media de las mediciones de cinco plantas diferentes con 3 réplicas técnicas cada uno de ellas. El descenso de la carga viral a 7 dpi en las plantas *atprimpol* es significativo según el T de Student ($p < 0,01$). También se indican el ajuste de los estándares a la recta patrón (R^2), la eficiencia de la reacción (E) y la ecuación de la recta patrón (Y).

3. Caracterización de las cuasiespecies crecidas en ausencia de DNA polimerasas TLS

Como una primera aproximación a la caracterización genética de las cuasiespecies virales de TYLCV en los fondos mutantes de DNA Pol TLS, seleccionamos una réplica biológica con una carga viral media de las líneas WT, *atrev1-1*, *atrev1-2*, *atpolh* y *atpolk* a 7 dpi. Estas muestras fueron amplificadas por RCA y secuenciadas por NGS en la plataforma Illumina HiSeq a partir de librerías de 300 nt, obteniéndose lecturas pareadas de 100 nt no solapantes.

Las lecturas secuenciadas fueron analizadas por medio de QuasiFlow. Todas las lecturas de menos de 100 nucleótidos y con una calidad inferior a QC30 fueron descartadas para el análisis. La **Tabla 3** muestra el resumen de los resultados obtenidos a partir del análisis individual de las 10 infecciones secuenciadas.

Tabla 3. Resumen de los resultados obtenidos de la secuenciación de TYLCV por HiSeq

	Línea mutante	Reads [x10 ⁷]	Nt. [x10 ⁹]	Coverage [x10 ⁵]	SNPs	Transiciones	Trans.	Ts/Tv	InDels
	WT	1,85	1,86	6,53	649	482	167	2,89	17
	<i>atrev1-1</i>	2,08	2,09	7,35	476	421	55	7,65	22
7d	<i>atrev1-2</i>	2,22	2,23	7,84	536	438	98	4,47	20
	<i>atpolh</i>	0,57	0,58	2,05	833	666	167	3,99	18
	<i>atpolk</i>	0,64	0,64	2,28	874	628	246	2,55	19

Tras aplicar el filtrado de lecturas de secuencias, se obtuvieron entre $5,7 \times 10^6$ y $2,2 \times 10^7$ lecturas pareadas de TYLCV, lo que correspondió a entre $6,4 \times 10^9$ y $2,23 \times 10^9$ nucleótidos secuenciados por muestra y un *coverage* de 5×10^5 nucleótidos por posición nucleotídica. Ninguna de las muestras analizadas presentó un *coverage* con valores inferiores a $2,05 \times 10^5$ nucleótidos por posición nucleotídica.

A continuación analizamos el número de posiciones nucleotídicas mutadas en las lecturas del virus con respecto a la referencia de TYLCV (GenBank: AF071228). Los resultados revelaron un descenso del número de posiciones nucleotídicas mutadas en las secuencias de TYLCV en las líneas *atrev1-1* y *atrev1-2* (con 476 y 536 posiciones nucleotídicas mutadas respectivamente), con respecto al control (649 posiciones mutadas).

Para averiguar la causa del descenso del número de mutaciones observado en las líneas *atrev1-1* y *atrev1-2*, determinamos el número de transiciones (ts) y transversiones (tv), así como su ratio. AtRev1 se trata de una TLS no procesiva con afinidad por sitios abásicos en los que introduce una C de manera preferente. Esto se traduce en una mayor introducción de transversiones G→C y A→C que de transiciones T→C. El ratio ts/tv en la infección del WT fue de 2,89, mientras que en las infecciones de las líneas *atrev1-1* y *atrev1-2* alcanzó valores de 7,65 y 4,47 respectivamente. Este aumento del ratio ts/tv en ausencia de ATREV1 se debió a un descenso drástico en las transversiones, como se muestra en la **Tabla 3**, mientras que las transiciones permanecieron a niveles similares al WT. También se observó un incremento del ratio ts/tv en la cuasiespecie con fondo *atpolh* (ts/tv=3,99) aunque más moderado que para los mutantes de ATREV1. Sin

embargo, en este caso la causa fue debida a un incremento en el número de transiciones mientras que el número de transversiones permaneció a niveles del WT.

En el caso de las infecciones de TYLCV en las líneas *atpolh* y *atpolk* no se observó descenso alguno en el número de las posiciones nucleotídicas mutadas de las secuencias virales. Esto podría verse justificado por el efecto de complementación que tienen estas polimerasas o bien a que *atpolh* es un mutante heterocigoto para *ATPOLH* y *atpolk* es un homocigoto *knockdown*.

En cuanto al número de mutaciones de tipo InDels, no se encontraron diferencias significativas entre el WT y ninguna de las muestras procedentes de infecciones sobre líneas mutantes según el ANOVA realizado por QuasiFlow.

Dado que en ausencia de AtRev1 disminuyó el número de posiciones mutadas en TYLCV, decidimos estudiar el papel de esta polimerasa de TLS en la variabilidad de TYLCV. Para ello realizamos un análisis más profundo de las cuasiespecies virales en infecciones de TYLCV en plantas *atrev1-1* y *atrev1-2*. Igualmente realizamos una caracterización de las cuasiespecies de TYLCV en plantas deficientes en PrimPol por su efecto en la acumulación viral a días tempranos.

Con este objetivo seleccionamos al azar dos muestras de las líneas WT y *atrev1-2*, *atrev1-1*, *atprimpol* a 7 dpi y tres muestras de cada una de ellas a 28 dpi. Las muestras se amplificaron por RCA y se secuenciaron por NGS mediante Illumina MiSeq con genotecas de 550 nt, obteniendo lecturas pareadas de 300 nt. El cambio de tecnología vino justificado por la mejora en la reconstrucción de haplotipos que ofrecen las lecturas de 300 nt con respecto a las de 100 nt tal como se ha explicado en detalle en el Capítulo 2.

El número de lecturas de calidad superior a QC30 y de tamaño mayor a 100 nt en todas las muestras analizadas estuvo entre $0,81 \times 10^6$ y $1,48 \times 10^6$ (Tabla 4), a excepción de las dos secuenciaciones de las infecciones en las líneas *atprimpol* a 7 dpi de la que se obtuvieron $2,5 \times 10^5$ y $4,4 \times 10^5$.

Tabla 4. Resumen de los resultados obtenidos por Quasiflow tras la secuenciación MiSeq de TYLCV

Tiempo	Fondo mutante	Nº muestra	Reads [x10 ⁶]	Nt [x10 ⁸]	Coverage [x10 ⁴]	Frecuencia mutación ^a [x10 ⁻³]	Nº haplotipos	Índice Shannon ^b	Diversidad ^c [x10 ⁻⁴]
7d	WT	1	1,41	2,42	8,20	1,66	1247	0,91	7,55
		2	1,36	2,35	8,01	1,66	1353	0,91	9,36
	<i>atrev1-1</i>	1	0,75	1,26	4,23	1,74	492	0,89	0,10
		2	1,06	1,80	6,13	1,65	1209	0,90	7,85
	<i>atrev1-2</i>	1	1,07	1,85	6,29	1,65	1121	0,89	7,43
		2	0,81	1,36	4,63	1,71	703	0,89	7,23
	<i>atprimpol</i>	1	0,25	0,41	1,39	1,76	310	0,88	6,95
		2	0,44	0,75	2,50	1,71	567	0,87	7,19
28d	WT	1	1,45	2,50	7,42	1,77	827	0,76	7,41
		2	0,93	1,57	4,49	1,94	372	0,71	0,11
		3	0,88	1,40	4,21	1,98	378	0,81	8,48
	<i>atrev1-1</i>	1	0,25	0,42	1,39	1,77	100	0,64	7,08
		2	1,15	1,86	5,57	1,85	440	0,79	8,59
		3	1,36	2,35	6,93	1,75	643	0,82	8,03
	<i>atrev1-2</i>	1	1,48	2,49	7,66	1,72	889	0,80	8,67
		2	1,19	2,00	6,06	1,79	518	0,81	8,24
		3	1,23	2,05	6,26	1,76	679	0,81	7,76
	<i>atprimpol</i>	1	1,15	2,05	5,97	1,77	645	0,80	7,23
		2	0,86	1,48	4,27	1,79	414	0,76	7,27
		3	0,81	1,46	4,39	1,72	674	0,76	7,36

^aLa frecuencia de mutación fue calculada como el sumatorio de las frecuencias de mutación para cada posición de la referencia.

^bEl índice de Shannon fue calculado a partir de las frecuencias obtenidas de cada haplotipo reconstruido para cada cuasiespecie, a partir de la formula descrita en Materiales y Métodos del capítulo 3.

Este descenso en el número de lecturas pudo deberse al menor número de secuencias de virus de partida en la amplificación por RCA. A pesar de ello, para ambas muestras se alcanzó el valor mínimo de *coverage* necesario, estimado en 10000X (**Figura 9**) por posición nucleotídica (ver justificación en el capítulo II) para poder incluirlas en el estudio comparativo con las otras muestras.

El *coverage* de las muestras osciló entre $1,39 \times 10^4$ de una de las muestras de *atprimpol* y $8,20 \times 10^4$ del control WT ambas a 7 dpi.

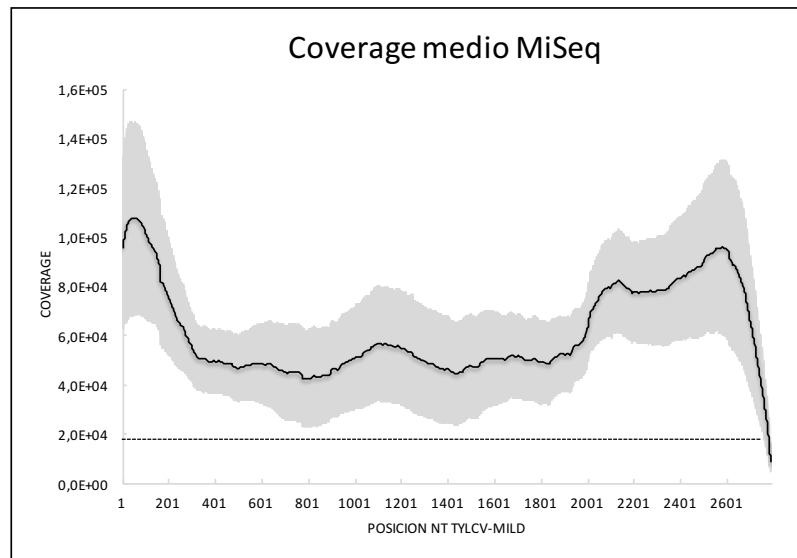


Figura 9. Cobertura media por nucleótido de las muestras secuenciadas por MiSeq. Se observó una cobertura mayor en los 200 primeros nucleótidos y una menor cobertura en los 100 últimos, debido a la naturaleza circular de las moléculas secuenciadas, QuasiFlow asoció las lecturas que compartían ambas regiones inicial y final a la primera de ellas, creando ese descenso en la región final. La línea discontinua representa la cobertura por nucleótido que QuasiFlow considera mínima para poder ofrecer resultados reproducibles entre réplicas biológicas (10000X).

La frecuencia de mutación de TYLCV no varió significativamente entre las diferentes líneas o el WT ni entre 7 y 28 dpi, aunque se observó un leve incremento en el WT entre los 7 dpi ($1,66 \times 10^{-3}$ mutaciones/nt de media) y los 28 dpi ($1,90 \times 10^{-3}$ mutaciones/nt de media). En cambio la frecuencia de mutación de TYLCV en las líneas *atprimpol*, *atrev1-1* y *atrev1-2* permaneció constante en torno a valores de $1,7 \times 10^{-3}$ mutaciones/nt. Estos valores de frecuencia de mutación resultaron ser hasta de un orden de magnitud superior a lo que tradicionalmente se había observado al analizar cuasiespecies de virus de ssDNA por secuenciación Sanger (Duffy and Holmes, 2008) pero coinciden con los valores mostrados en (Richter et al., 2016a).

Se realizó una reconstrucción de haplotipos con Quasiflow y se calculó la frecuencia de aparición de estos haplotipos en la cuasiespecie, índice de Shannon y la diversidad nucleotídica.

El índice de Shannon fue similar en las cuasiespecies de TYLCV en las plantas *atrev1-1*, *atrev1-2* y *atprimpol* a 7 dpi con valores entre 0,87 y 0,91, no mostrando diferencias significativas con respecto al control. A 28 dpi el índice de Shannon tampoco presentó

diferencias significativas de ninguna cuasiespecie con respecto al control, pero sí se observó un descenso de este índice con valores entre 0,71 y 0,81.

4. Caracterización del espacio de secuencia y variabilidad genética de las cuasiespecies virales de TYLCV en ausencia de REV1 y PRIMPOL

Para determinar si la ausencia de ATREV1 y PRIMPOL puede modificar el espacio de secuencia que exploran las cuasiespecies, analizamos el espectro de mutantes de las 20 muestras secuenciadas por NGS MiSeq (8 muestras a 7dpi –2 WT, 2 *atrev1-1*, 2 *atrev1-2* y 2 *atprimpol-* y 12 muestras a 28 dpi –3 WT, 3 *atrev1-1*, 3 *atrev1-2* y 3 *atprimpol-*). Mediante la herramienta QuasiFlow estudiamos el número de posiciones mutadas con respecto a la referencia en cada muestra, su frecuencia y la distribución de las mutaciones en el genoma de TYLCV, tanto en nucleótidos como en codones en los 6 ORFs. También realizamos un estudio pormenorizado de los Indels en cada muestra, analizándolos de forma individual y conjunta y por último se llevó a cabo el estudio de recombinaciones halladas en cada una de las muestras, diferenciando si éstas ocurrían entre cadenas de la misma polaridad o entre cadenas de polaridad opuesta. La Figura 10 muestra una representación gráfica de los análisis realizados por QuasiFlow con todos los cambios posicionales y la distribución de mutaciones por nt, por codón y recombinaciones para cada muestra. A modo de ejemplo, en la figura 10 se muestran los circos generados por QuasiFlow de una cuasiespecie analizada a 7dpi sobre fondo WT y otra analizada a 30 dpi. El resto de representaciones pueden verse en el Anexo Circos TYLCV.

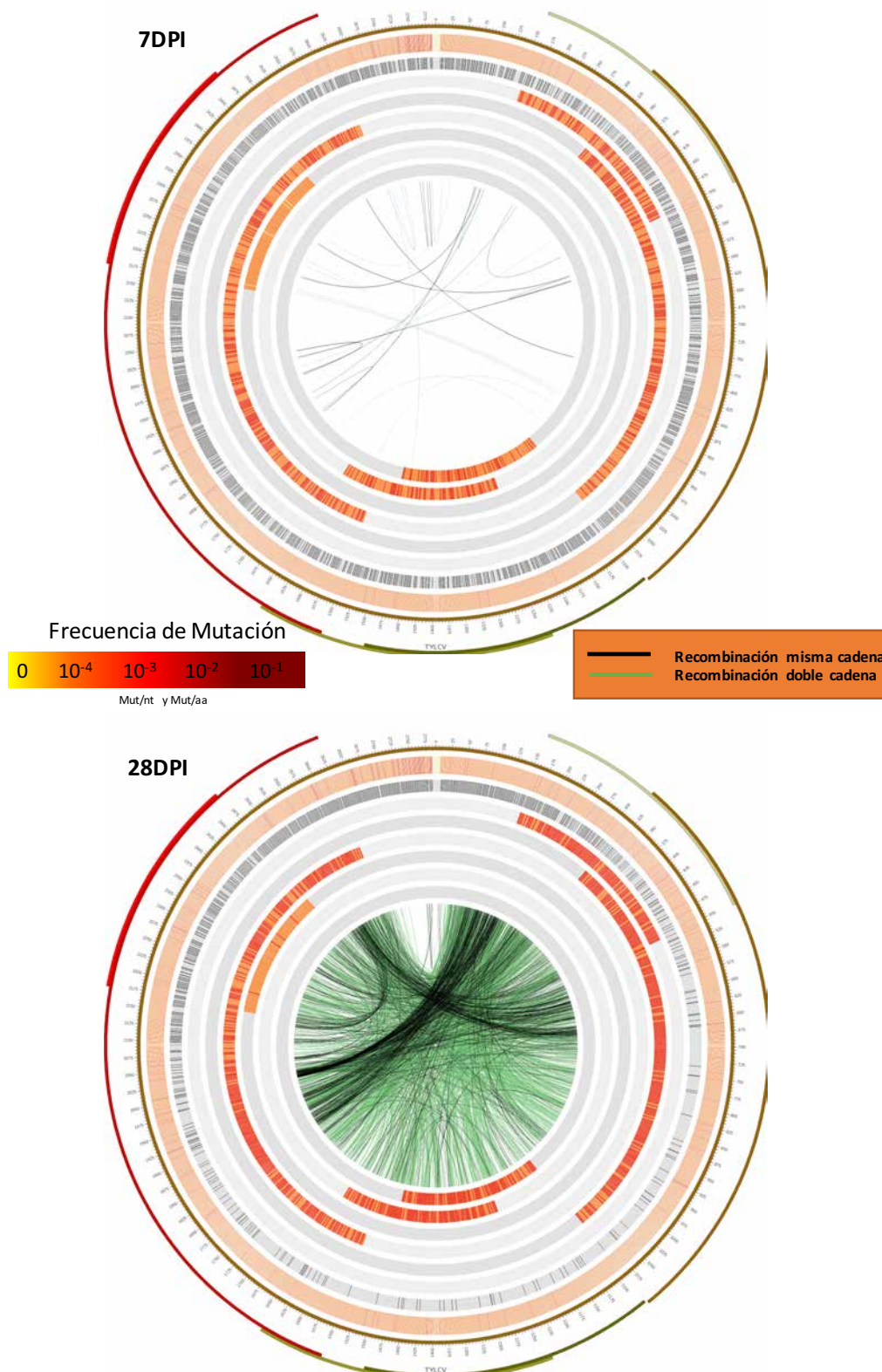


Figura 6. Representación gráfica del perfil de mutaciones de las poblaciones de TYLCV-Mild a 7 y 28dpi. La frecuencia de mutación de nucleótidos y de codones está expresada por un código de colores del amarillo al rojo. Y la representación de puntos de recombinación, diferenciando entre recombinaciones entre cadenas de la misma polaridad (línea verde) o de polaridades distintas (línea negra).

4.1. Análisis de mutaciones de TYLCV en ausencia de Rev1 y PrimPol

La distribución de posiciones mutadas con respecto a la referencia de TYLCV-Mild fue uniforme a lo largo de todo el genoma vírico en todas las muestras analizadas. No se encontró ninguna diferencia entre las cuasiespecies control y las de las líneas mutantes, ni a 7dpi, ni a 28 dpi tal como se observa en la **Figura 11**. Además, QuasiFlow generó una tabla con todas las posiciones mutadas de cada muestra, marcando el tipo de cambio de nucleótido y su frecuencia, que pueden consultarse en el **Anexo SNPs TYLCV**. A los 7 dpi las cuasiespecies de TYLCV en los fondos mutantes, *atrev1-1*, *atrev1-2* y *atprimpol* acumularon un menor número de posiciones mutadas con respecto al control. Este efecto del fondo mutante desaparece a los 28 días. También se observó un descenso de posiciones nucleotídicas mutadas en el WT entre los 7 y los 28 dpi, de 1300 a unas 900 posiciones mutadas, mientras que en el caso de las cuasiespecies que infectaron plantas

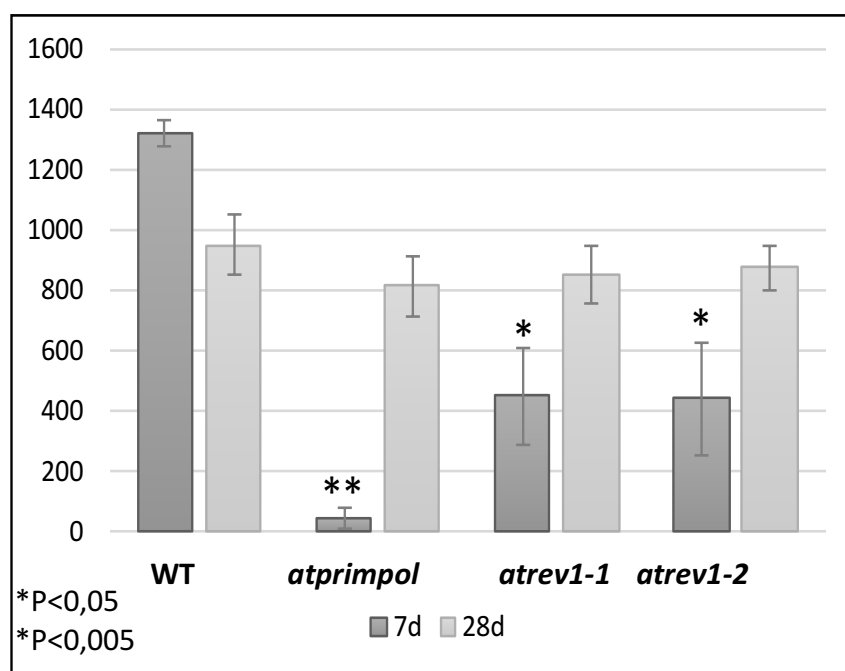


Figura 11. Posiciones nucleotídicas variables en las infecciones de TYLCV de los diferentes fondos mutantes de TLS. Plantas de las líneas WT y mutantes para las DNA polimerasas AtPrimPol (*atprimpol*) y AtRev1 (*atrev1-1* y *atrev1-2*) se agroinocularon con el begomovirus TYLCV. A los 7 y 28 dpi se recogieron las plantas, se extrajo el DNA que se amplificó por RCA y secuenció por Illumina MiSeq. Las lecturas pareadas de 300 pb se analizaron con QuasiFlow. Las barras de error representan la desviación estándar entre réplicas biológicas.

atprimpol, *atrev1-1* y *atrev1-2* el efecto es el inverso, ya que se pudo observar un incremento importante de posiciones mutadas que llegan a niveles del WT a los 28 dpi.

Estos resultados hay que tomarlos con precaución puesto que un análisis posterior indicó una relación directa entre el número de estas posiciones mutadas y el número de lecturas secuenciadas ($p < 0,01$) en cada muestra tal como se muestra en la **Figura 12**. Aunque los valores de R^2 de la recta quedaron muy por debajo de los valores recomendables ($R^2 = 0,803$), para evitar posibles sesgos derivados del número de secuencias de cada muestra decidimos analizar el tipo de cambios que ocurrían en las cuasiespecies de TYLCV en las plantas control y mutantes. Para confirmar que los descensos de posiciones mutadas se debían, al menos en parte, al efecto de la ausencia de AtPrimPol y AtRev1 se estudiaron los tipos de mutaciones de cada una de las muestras a 7dpi y a 28 dpi relativizándolas al WT para cada tiempo estudiado.

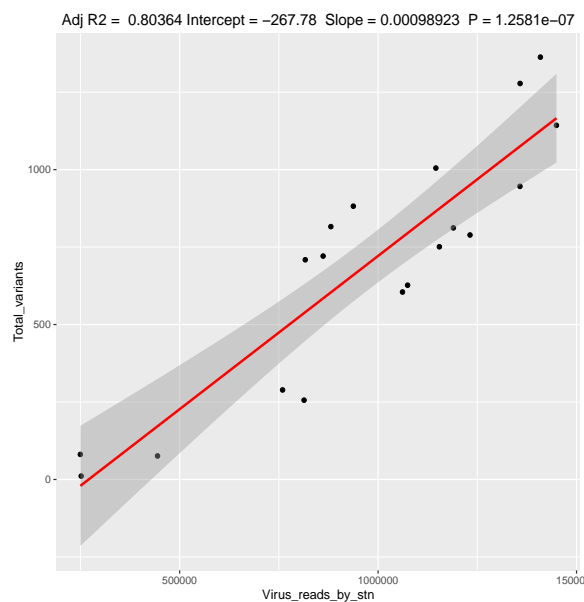


Figura 12. Relación entre número de lecturas y variantes. Gráfica extraída de QuasiFlow donde se muestra la relación directa entre el número de lecturas en cada secuenciación (Virus_reads_by_stn) y el número de variantes encontrados en ellas (Total_variants). Cada punto representa una de las 20 muestras analizadas por NGS.

Los análisis de las muestras de mutantes *atrev1* obtenidas a 7dpi (**Figura 13**) revelan tanto en las infecciones de TYLCV en *atrev1-1* como en las de *atrev1-2* un descenso general de todos los tipos de mutaciones, siendo la mayor disminución de mutaciones en los tipos A→T, T→A, A→G, T→G, C→G, A→C, T→C, G→C, con valores entre 0,0 y 0,4 veces el WT. Estas mutaciones, excepto A→T y T→A, pueden asociarse a las ocasionadas

por la acción de AtRev1 al insertar una C en sitios abásicos. Este mismo efecto se mantiene a 28 dpi donde nuevamente se observa un descenso de mutaciones tipo T→G, C→G y G→C con valores de 0,1, 0,2 y 0,5 veces el WT, respectivamente. Nótese la reproducibilidad de los resultados en los mutantes *atrev1-1* y *atrev1-2* a 7 y 28 dpi, que como se observa en la **Figura 13** son iguales entre sí.

En las cuasiespecies de TYLCV en las plantas *atprimpol* a 7 dpi detectamos un descenso generalizado de todos los tipos de mutaciones, con valores entre 0,0 y 0,5 excepto en las mutaciones C→G con un valor de 0,8 veces el control. A 28 dpi, en cambio, para las mutaciones T→G dicha reducción fue de tan sólo 0,4 veces a las halladas en el WT.

7dpi	<i>atrev1-2</i>			
↷	A	T	G	C
A		0,2	0,3	0,4
T	0,2		0,3	0,4
G	0,5	0,6		0,4
C	0,5	0,7	0,0	

7dpi	<i>atrev1-1</i>			
↷	A	T	G	C
A		0,2	0,3	0,5
T	0,2		0,4	0,3
G	0,5	0,6		0,1
C	0,5	0,9	0,0	

7dpi	<i>atprimpol</i>			
↷	A	T	G	C
A		0,1	0,0	0,1
T	0,1		0,1	0,1
G	0,1	0,1		0,5
C	0,5	0,2	0,8	

28dpi	<i>atrev1-2</i>			
↷	A	T	G	C
A		0,7	1,2	0,5
T	0,5		0,1	1,0
G	1,1	0,5		0,2
C	0,7	1,0	0,2	

28dpi	<i>atrev1-1</i>			
↷	A	T	G	C
A		0,8	1,0	0,6
T	0,7		0,1	1,0
G	0,9	0,8		0,5
C	0,7	0,9	0,3	

28dpi	<i>atprimpol</i>			
↷	A	T	G	C
A		1,3	1,1	0,8
T	0,9		0,4	0,9
G	1,0	1,0		0,7
C	1,1	1,0	1,2	

Figura 13. Mutaciones encontradas en las secuencias de TYLCV en las infecciones de *atprimpol*, *atrev1-1* y *atrev1-2*. Los números están relativizados al total de mutaciones encontradas en el fondo WT a 7 y 28 dpi respectivamente. Los valores inferiores a 1 indican un descenso de esas mutaciones con respecto al WT, mientras que los valores superiores a 1 indicarían un incremento de las mismas con respecto al WT

A continuación examinamos el ratio Ts/Tv que, al contrario que el número de posiciones mutadas, no tiene relación directa con el número de lecturas de cada muestra, ni con la frecuencia de mutación de las cuasiespecies ($p > 0,01$, $R^2 < 0,5$). Los valores Ts/Tv en las infecciones control fueron de 1,90 y 1,2 a 7 y 28 dpi, respectivamente, tal como se muestran en la **Figura 14**, que son los valores esperados para poblaciones virales, entre 1 y 2. En ausencia de AtPrimPol los valores de Ts/Tv de las cuasiespecies se situaron a niveles del WT tanto a 7 dpi como a 28 dpi. Sin embargo, en las dos cuasiespecies de

TYLCV en ausencia de AtRev1 el ratio Ts/Tv aumentó hasta más del doble del observado en el WT a 7 dpi llegando a valores de 3,4 y 3,5 de media para *atrev1-2* y de 2 en *atrev1-1*, respectivamente. Estos valores continuaron siendo superiores a los encontrados en el WT a los 28 dpi, alcanzando valores de 2 en *atrev1-1* y de 3 en *atrev1-2* y, mientras que el WT fue de 1,2.

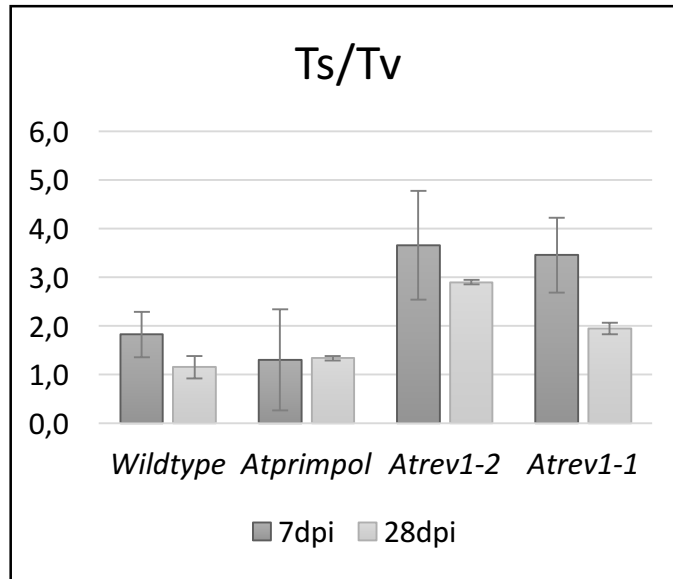


Figura 14. Ratio Ts/Tv de las mutaciones encontradas en las cuasiespecies de TYLCV en *A. thaliana* WT y mutantes *atprimpol*, *atrev1-1* y *atrev1-2* a 7 y 28 dpi. Las barras de error indica la desviación estándar para dos muestras a 7dpi y para 3 muestras a 28dpi.

Para averiguar cómo afectaba la ausencia de PrimPol y Rev1 a TYLCV en los codones de los ORFs de su genoma, se analizó la frecuencia de mutación por codón en los ORFs tanto no solapantes como solapantes. Como se observa en la **Figura 15**, a 7dpi no se observaron diferencias significativas en la frecuencia de mutación por codón entre las cuasiespecies de TYLCV control y en ausencia de PrimPol y Rev1, siendo para los ORFs V2, CP, Rep, REn y TrAP del orden de 4×10^{-3} mutaciones/codón. Sin embargo, el ORF codificante de C4 que es completamente solapante con el extremo 5' del ORF de REP tuvo una frecuencia de mutación mucho menor que los otros 5 ORFs (entre $5,61 \times 10^{-4}$ y $6,28 \times 10^{-4}$ mutaciones/codón, a excepción de una réplica sobre fondo *atprimpol* que ascendió a $6,76 \times 10^{-3}$ mutaciones/codón). El de las muestras a los 28 días mostró un ligero aumento en la frecuencia de mutación de los ORFs CP, Rep, REn y TrAP que fue

del orden de 6×10^{-3} mutaciones/codón, mientras que en V2 aumentó hasta $7,0 \times 10^{-3}$ mutaciones/codón. De nuevo C4 mantuvo los niveles más bajos de frecuencia de mutación a 28 dpi (entre $9,0 \times 10^{-4}$ y $3,0 \times 10^{-3}$ mutaciones/codón).



Figura 15. Representación gráfica de la frecuencia de mutación por codón de los 6 ORFs de TYLCV crecidos sobre los fondos mutantes *atprimpol*, *atrev1-1*, *atrev1-2* y WT. Cada vértice del hexágono representa un ORF, la frecuencia de mutación es mayor cuanto más se aleja el valor del centro del gráfico. A 7 dpi se observa cierta homogeneidad en la frecuencia de mutación de todos los ORFs excepto de la C4, cuya frecuencia de mutación es inferior en todas las cuasiespecies de TYLCV, independientemente del fondo mutante. A 28 dpi se incrementa la frecuencia de mutación de los ORFs pero se sigue manteniendo la baja frecuencia de mutación del ORF C4 con respecto al resto.

4.2. Análisis de InDels de TYLCV en ausencia de Rev1 y PrimPol

Otra fuente de variabilidad importante en los virus son las inserciones y deleciones (Domingo et al., 2012). Por esto, hemos analizado tanto el número total de InDels por muestra, así como su posición y su frecuencia. La tabla de posiciones y frecuencia se puede consultar en el **Anexo InDels TYLCV**.

Tradicionalmente se ha considerado que la pérdida de una o más bases en la secuencia de un genoma es un evento que puede ocurrir más frecuentemente que las adiciones de bases o inserciones. En nuestro estudio, el número de inserciones que se observó en cada muestra fue siempre muy inferior al de deleciones. En el 90% de los InDels su frecuencia no superó el 0,3%, y en ningún caso el 1%. Tan sólo hubo una deleción con una frecuencia por encima de 0,8%. Sin embargo, resultó llamativo que muchas de las deleciones fuesen halladas en todas las cuasiespecies analizadas (**Figura 16**), encontrándolas tanto a 7 dpi como a 28 dpi. Es el caso de las deleciones en las posiciones

163, 167, 661, 923 1654, 1699, 1799, 2074, 2502, 2527 y 2627 a 7dpi que se encuentran en al menos 7 de las 8 poblaciones virales analizadas. Y las posiciones 661, 923, 1699, 1799, 1935, 2074, 2251, 2502, 2527, 2590, 2697, 2656 y 2722 que fueron halladas en al menos 11 de las 12 muestras analizadas a 28 dpi. Además tal como se observa en la **Figura 16**, existen otras muchas deleciones compartidas por más de una muestra (33 a 7 dpi y 51 a 28 dpi) lo que sugiere que se trata de una “zona caliente” de mutación con una mayor predisposición de esta secuencia a generar deleciones en esas posiciones. Sorprendentemente, la región comprendida entre el nt 1000 y el 1600 apenas contienen alguna de estas “deleciones recurrentes”.

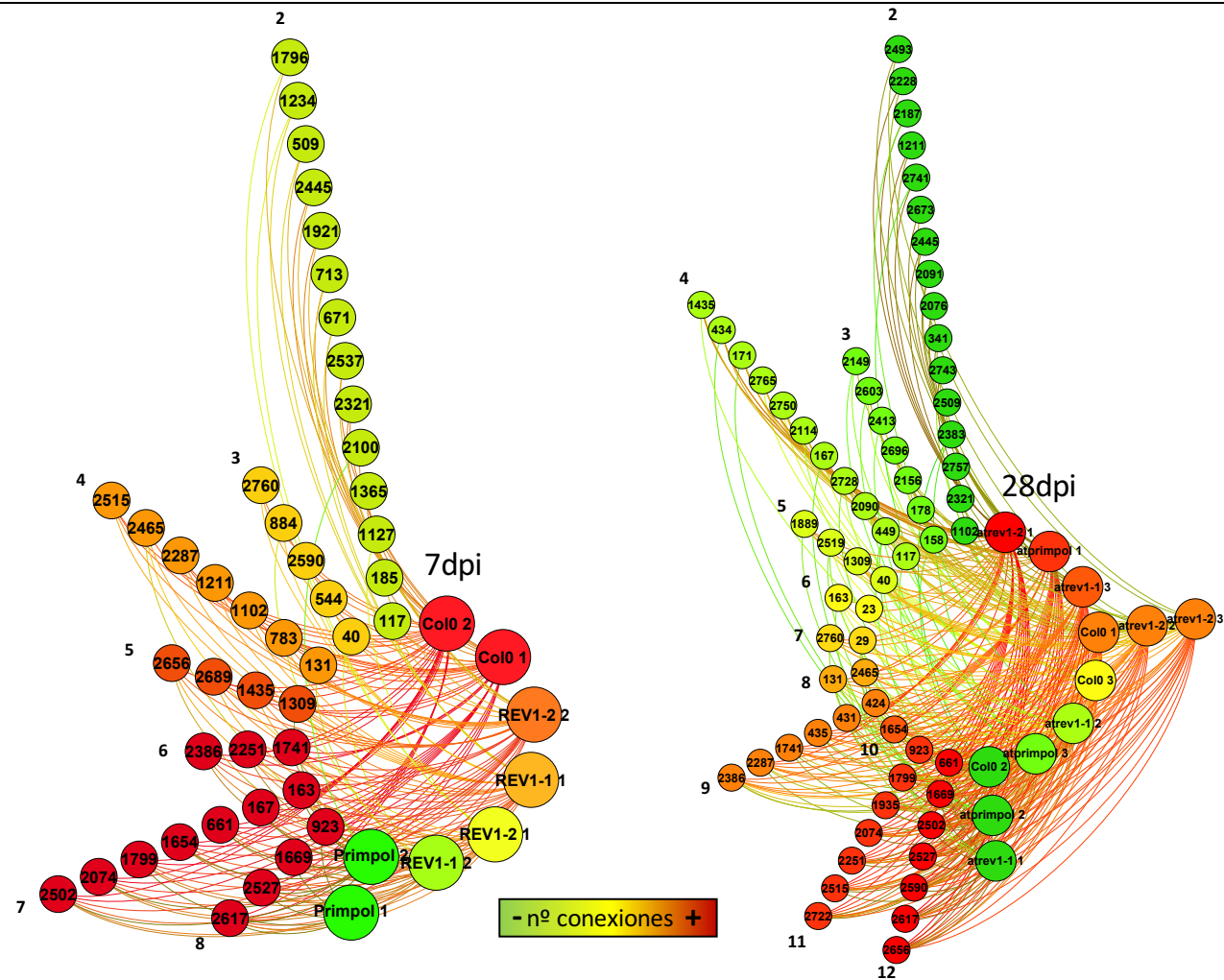


Figura 16. Red de interacciones de las deleciones compartidas entre cuasiespecies de TYLCV infectando *A. thaliana* Col-0 WT y mutantes de TLS y de PrimPol. Se muestran los eventos de deleción (círculo con la posición nucleotídica) que comparten dos o más de estas cuasiespecies (círculos más grandes con el nombre de la cuasiespecie). Los colores varían en función del número de conexiones que llegan a ese nodo, de forma que una deleción que esté en todas las cuasiespecies tendrá color rojo, al igual que lo tendrá una cuasiespecie con muchas deleciones. El número en los extremos de las líneas de los nodos de las deleciones indica el número de cuasiespecies que comparten esas deleciones.

4.3. Análisis de eventos de recombinación en las cuasiespecies virales de TYLCV en ausencia de Rev1 y PrimPol

El estudio de los eventos de recombinación en las poblaciones virales puede proporcionar una valiosa información acerca de cómo los virus emergentes reorganizan su genoma, generando con ello nuevas variantes genéticas. Estas recombinaciones pueden ser de dos tipos dependiendo de las cadenas de DNA del virus estén implicadas: recombinaciones entre cadenas de la misma polaridad o recombinaciones entre cadenas de polaridad contraria. Estas últimas requieren dos eventos de rotura de la doble cadena para que puedan originarse. En una primera aproximación se analizó el número de secuencias recombinantes totales encontradas en cada una de las 20 muestras secuenciadas, las 8 cuasiespecies a 7dpi y las 12 a 28 dpi (**Tabla 5**).

Tabla 5. Número de eventos de recombinación hallados en TYLCV infectando *A. thaliana* WT y mutantes de DNA Pol TLS a 7 dpi y 28 dpi

		Neg. a pos. Pos. a neg.	Pos. a pos. Neg. a neg.
7d	<i>Wild type</i>	1412,5 ± 347	1553,5 ± 136
	<i>atrev1-2</i>	855,5 ± 91	989 ± 92
	<i>atrev1-1</i>	1126,5 ± 402	780 ± 264
	<i>atprimpol</i>	726 ± 229	448 ± 102
28d	<i>Wild type</i>	99860 ± 41975	33670,6 ± 8655
	<i>atrev1-2</i>	84984 ± 10574	37856,6 ± 1260
	<i>atrev1-1</i>	107521,5 ± 28834	28377,3 ± 4248
	<i>atprimpol</i>	78594,6 ± 21396	38133,6 ± 9405

A 7 dpi observamos en todas las infecciones el mismo número de eventos de recombinación entre cadenas directas que dobles. Las cuasiespecies con mayor número de eventos de recombinación de ambos tipos pertenecieron a las infecciones de plantas WT, con 1412,5 ± 347 recombinaciones directas y 1553,5 ± 136 recombinaciones de doble cadena. Sin embargo, en las infecciones en ausencia de AtRev1 el número máximo de recombinaciones directas de TYLCV ocurrió en el mutante *atrev1-1* (1126,5 ± 402), y el de recombinaciones de doble cadena fue de 989±92 en *atrev1-2*, muy por debajo de las encontradas en la infección con fondo WT. Todas las recombinaciones halladas en

las 20 muestras pueden verse en el **Anexo Recombinaciones TYLCV**. En el caso de las infecciones en ausencia de AtPrimPol pudo observarse un descenso aún mas significativo con respecto a la infección WT, ya que se detectaron 726 ± 229 recombinaciones directas y 448 ± 102 de doble cadena, 3 veces menos que en el control. Los resultados del análisis de las recombinaciones a 28 dpi difieren significativamente de los mostrados a 7 dpi ($p < 0,01$). En todos los casos se hallaron valores muy superiores a los observados a 7 dpi que superaron los $78594,6 \pm 21396$ eventos de recombinación de cadena directa y los $28377,3 \pm 4248$ para la doble. Cuando se comparó el número de recombinaciones directas en los fondos mutantes con respecto al WT no se observaron diferencias significativas. Por otra parte se constataron diferencias significativas ($p < 0,01$) en todas las muestras analizadas entre el número de recombinaciones directas y el de recombinaciones dobles, que en todos los casos fue de 3 veces menor en estas últimas.

También analizamos los eventos de recombinación mediante *scatter plots* para evaluar si existía algún cambio en el patrón de recombinaciones de TYLCV cuando se eliminaban Rev1 y PrimPol. Para ello se analizaron las recombinaciones que estuviesen repetidas un mínimo de 5 veces sin considerar las que ocurrían con menor probabilidad para evitar enmascarar los patrones de recombinación. A continuación se unieron todos los datos de las réplicas pertenecientes a los mismos fondos mutantes y se representaron en un eje bidimensional (**Figura 17**). Este estudio sólo se muestra a 28 dpi puesto que a 7 dpi no pudimos detectar ningún tipo de patrón. El patrón de recombinación entre cadenas de la misma polaridad es completamente diferente al encontrado entre cadenas de diferente polaridad. No se observaron diferencias en los patrones de recombinación de TYLCV entre los fondos WT y mutantes *atprimpol* y *atrev1*, en ninguno de los dos tipos de recombinaciones. Lo único destacable es una tendencia en el WT y el *atrev1* a que la región entre los nt 2000-2600 recombine con la región entre los nt 800-1100. Dicha tendencia no se observa en las cuasiespecies de TYLCV en fondo *atprimpol*, mostrando diferencias significativas entre la línea *atprimpol* y la línea WT ($P < 0,05$) según la T de Student aplicada para estas muestras, aunque son necesarios más estudios para confirmar esa tendencia.

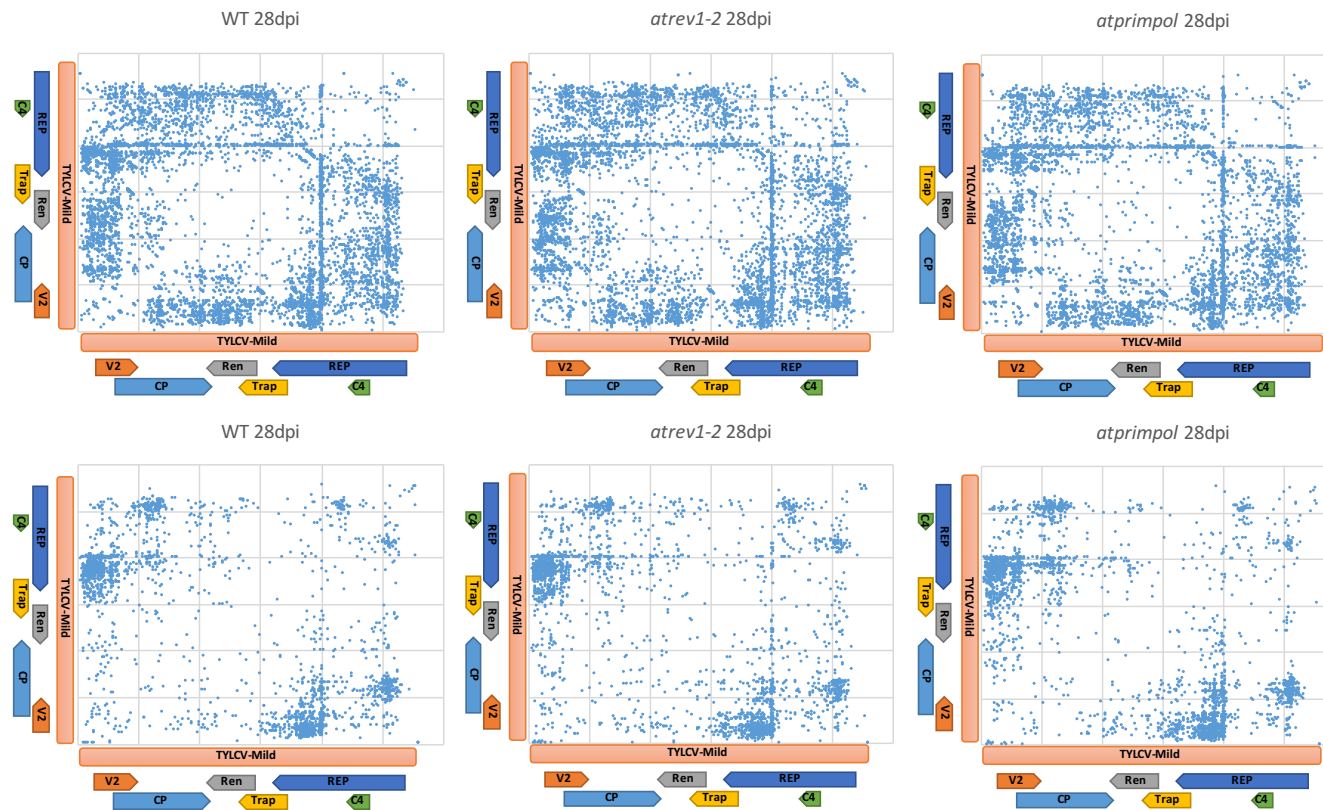


Figura 17. Scatter plot de las recombinaciones halladas en cuasiespecies de TYLCV infectando *A. thaliana* WT y mutantes *atprimpol* y *atrev1-2* a 28 dpi. En ambos ejes de las gráficas se representa el genoma de TYLCV y sus ORFs para facilitar la ubicación de los eventos de recombinación que ocurren dentro de la cuasiespecie. La gráfica superior muestra las recombinaciones entre cadenas de diferente polaridad, mientras que la grafica inferior indica las recombinaciones que ocurren entre cadenas con la misma polaridad. No se observaron diferencias en la distribución de los eventos de recombinación entre las cuasiespecies.

5. Reconstrucción de haplotipos en las cuasiespecies virales de TYLCV en ausencia de Rev1 y PrimPol

El análisis de los datos con QuasiFlow permitió la reconstrucción de haplotipos de cada una de las cuasiespecies. Tal como se explica en el capítulo II, el flujo de trabajo realiza un análisis iterativo de las lecturas, uniéndolas si son idénticas en un 90% de su longitud creando haplotipos. De esta forma, las lecturas ciclo a ciclo disminuyen en número, y aumentan en longitud de nucleótidos hasta generar todos los haplotipos posibles.

La decisión de utilizar la secuenciación MiSeq en vez de HiSeq vino justificada por la necesidad de optimizar este paso de reconstrucción de haplotipos. Con MiSeq obtuvimos lecturas pareadas de 300 nt a partir de genotecas de 550 nt, en vez de las lecturas pareadas de 100 nt de HiSeq. Así, con una mayoría de secuencias completas de 550 nt, es posible reconstruir haplotipos de 2791 nt con gran fiabilidad.

El número de haplotipos encontrados en cada muestra fue altamente variable, ya que cada infección evoluciona de forma diferente en cada hospedador (**Tabla 4**). No obstante, como norma general se pudo observar que a 7 dpi el número de haplotipos era significativamente superior que a 28 dpi en todas las líneas excepto en *atprimpol* cuyo número de haplotipos (310 y 567), resultó mucho menor que el control (1247 y 1353 haplotipos reconstruidos). Además, dado que QuasiFlow reveló que no existía relación directa entre el número de haplotipos reconstruidos y el número de lecturas de cada línea ($p > 0,01$) ese descenso de haplotipos en las líneas *atprimpol* a 7 dpi resulta significativo.

6. Reconstrucción de la cuasiespecie viral de TYLCV en ausencia de Rev1 y PrimPol

Con el objetivo de visualizar cada una de las cuasiespecies de una forma más intuitiva se creó una red que conectase cada haplotipo de la cuasiespecie con nodos que representasen las posiciones nucleotídicas mutadas (SNPs e InDels), y de esta forma representar su “huella digital”. Además esta representación permite calcular la frecuencia de los haplotipos dentro de la población viral. Sirva como ejemplo una muestra a 7 dpi en la **Figura 18** y otra a 28 del WT en la **Figura 19**. En el caso de la muestra a 7 dpi, y como sucede en la mayoría de cuasiespecies reconstruidas, podemos observar que los haplotipos (círculos verdes) son más numerosos que en la cuasiespecie control a 28 dpi. Otra característica de las cuasiespecies a 7 dpi es que los haplotipos están

vinculados en su mayoría a un solo SNP (círculo naranja) o a una delección (círculo amarillo), lo que provoca que haya pocos haplotipos relacionados entre sí. Sin embargo, a 28 dpi podemos observar en algunos haplotipos un aumento del número de SNPs que son compartidos con más haplotipos. También aparecen más delecciones a 28 dpi.

El resto de reconstrucciones pueden verse en el **Anexo Reconstrucción de Haplotipos TYLCV**.

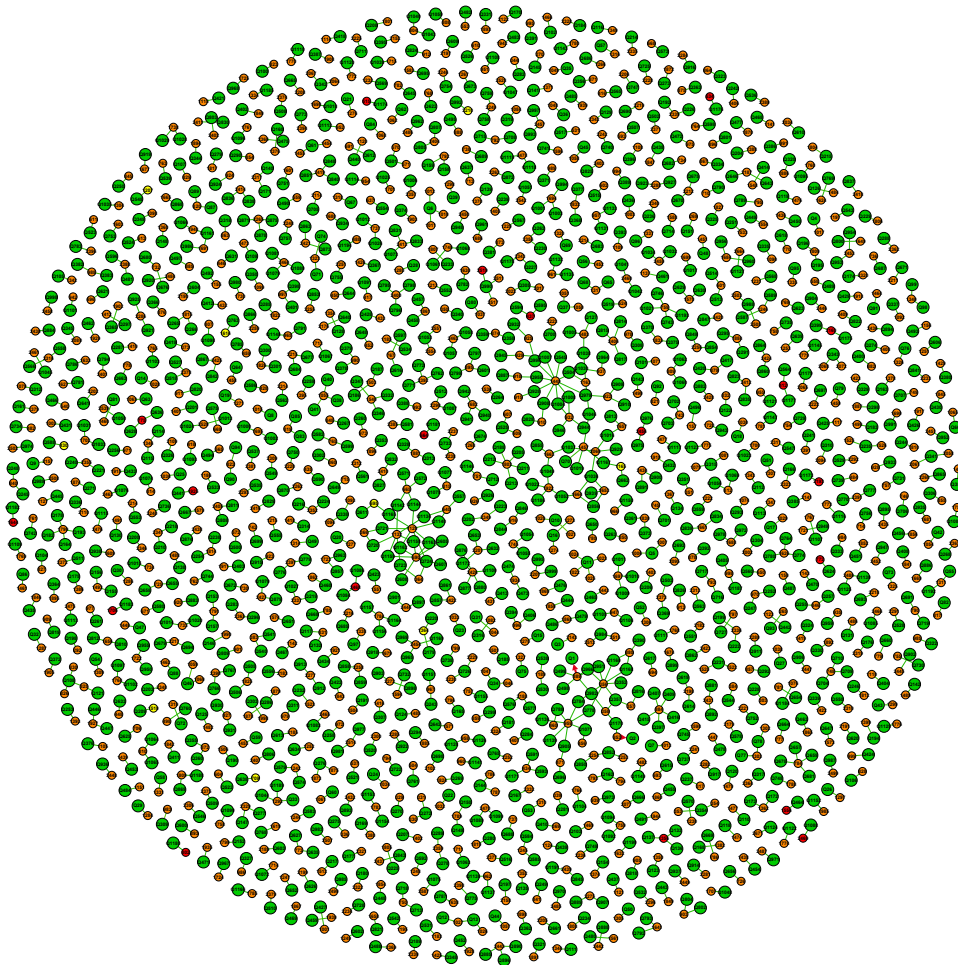


Figura 18 Representación gráfica de la distribución de haplotipos reconstruidos para la muestra 1 de TYLCV crecida sobre fondo WT a 7dpi. Los haplotipos se representan como los círculos de mayor tamaño con la nomenclatura “Q_n” y en color verde. Los cambios de nucleótidos con respecto a la secuencia de TYLCV-Mild de cada haplotipo presentan como círculos pequeños con un indicativo numérico que hace referencia a la posición con respecto a la referencia en la que se ha producido el cambio, los SNPs se representan con color naranja, las delecciones de color amarillo y las inserciones de color rojo. El grosor de las flechas indica la abundancia relativa de ese cambio dentro de la población viral.

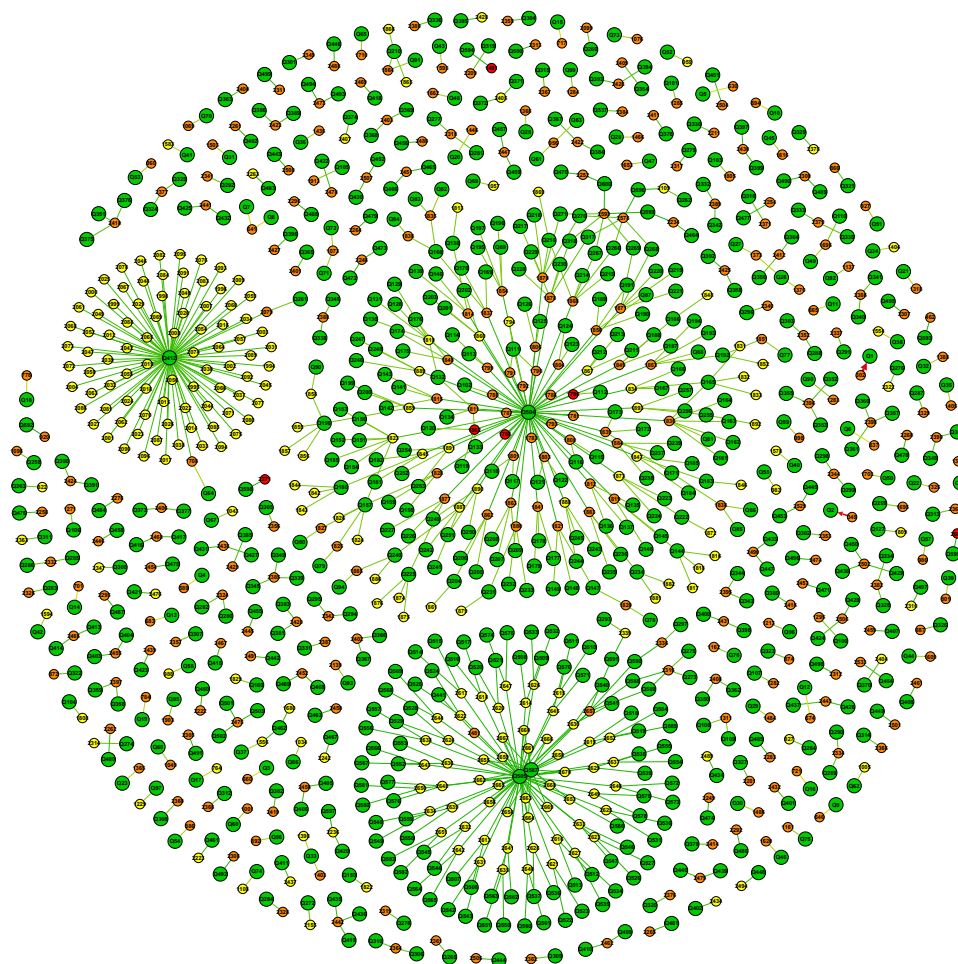


Figura 19 Representación gráfica de la distribución de haplotipos reconstruidos para la muestra 1 de TYLCV crecida sobre fondo WT a 28dpi. Los haplotipos se representan como los círculos de mayor tamaño con la nomenclatura “Q_n” y en color verde. Los cambios de nucleótidos con respecto a la secuencia de TYLCV-Mild de cada haplotipo presentan como círculos pequeños con un indicativo numérico que hace referencia a la posición con respecto a la referencia en la que se ha producido el cambio, los SNPs se representan con color naranja, las deleciones de color amarillo y las inserciones de color rojo. El grosor de las flechas indica la abundancia relativa de ese cambio dentro de la población viral.

7. Análisis multivariable de las infecciones de TYLCV sobre fondos genéticos mutantes de *Arabidopsis thaliana*

Al objeto de tener una visión global de la variabilidad de las cuasiespecies de TYLCV en ausencia de TLS y PrimPol, realizamos un análisis de componentes principales (PCA) de las 20 muestras secuenciadas por MiSeq para este estudio (8 muestras a 7dpi y 12 muestras a 28dpi) en el cual se incluyeron un total de 23 variables: Ts/Tv, índice de

Shannon, transiciones, transversiones, frecuencia de mutación de los seis ORFs, SNPs y tipos de SNPs relativos al número de lecturas para cada muestra.

QuasiFlow realizó un mapeado de las variables como se observa en la **Figura 20**, en la cual, el eje X o dimensión 1 (Dim1) justifica el 53,15% de la variabilidad entre las muestras, mientras que el eje Y o dimensión 2 (Dim2) justifica el 26,23% de la variabilidad.

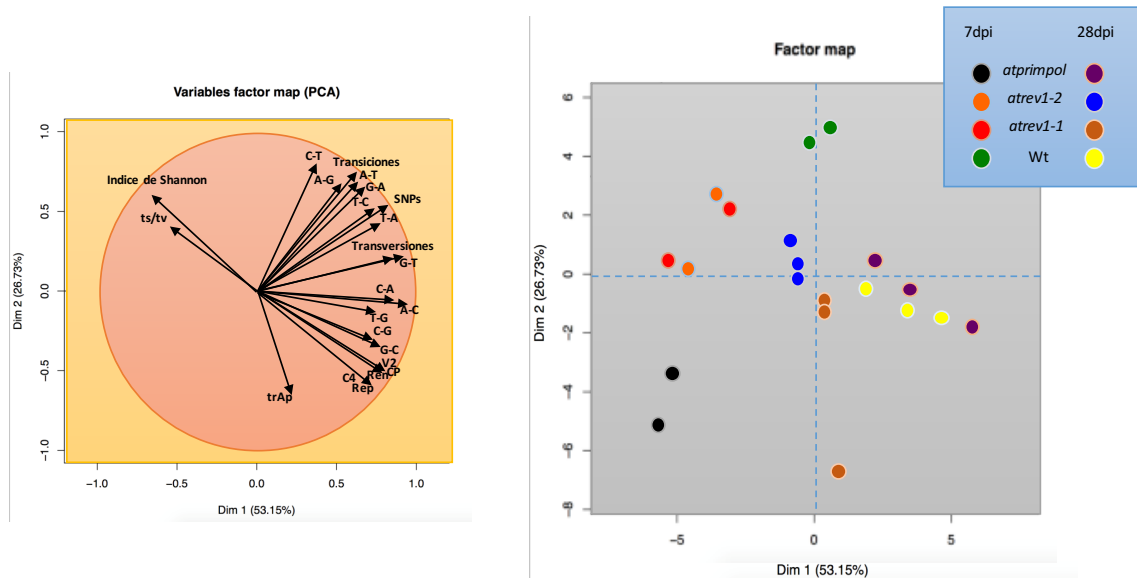


Figura 20. Análisis de componentes principales (PCA) para las muestras de TYLCV. A la izquierda se puede observar la distribución de las 23 variables sujetas a estudio para cada muestra. A la derecha se puede observar la distribución y agrupación de cada una de las muestras analizadas por PCA (puntos).

La representación de las muestras en el PCA muestra una clara agrupación por réplicas de todas las muestras, tanto a 7 dpi como a 28 dpi, indicando que el sistema es muy robusto y permite estudiar cuasiespecies virales a partir de réplicas biológicas. Además el PCA muestra que a 7 dpi las cuasiespecies crecidas en ausencia de Rev1 y PrimPol presentan diferencias muy importantes con respecto al WT, tanto en el la Dim1 como en la Dim2.

A los 28 dpi las diferencias justificadas por la Dim2 desaparecen, pero las cuasiespecies virales crecidas en fondos mutantes *atrev1-1* y *atrev1-2* siguen claramente diferenciadas del WT mientras que las que crecieron en ausencia de PRIMPOL se comportan como WT, sugiriendo que a 28 dpi no hay efecto del mutante *atprimpol* sobre la cuasiespecie, mientras que Rev1 sigue teniendo efecto sobre la población viral.

Discusión

La variabilidad genética de las poblaciones víricas refleja la complejidad de las interacciones que ocurren entre el virus y su huésped durante la infección. La mayoría de virus ssDNA utilizan el complejo de replicación del huésped que parasitan y posiblemente el daño producido en su genoma sea también reparado por la propia maquinaria de reparación celular. Como ya se ha comentado en la Introducción del presente capítulo, las DNA Pol TLS pertenecientes a la familia Y tienen un centro activo más abierto que las DNA polimerasas replicativas y pueden acomodar las lesiones presentes en el DNA y, por tanto, sintetizar a través de éstas, con una baja fidelidad de copia. En esta Tesis planteamos la hipótesis de que las DNA Pol TLS podrían actuar en la reparación de daños producidos en el genoma de virus ssDNA y, por consiguiente, incrementando el número de mutaciones de los genomas víricos.

Existen trabajos previos que apoyan la idea de que un virus durante su infección puede interactuar y alterar el mecanismo de reparación de DNA asociado a las DNA Pol TLS. En el caso del *Epstein Bar-virus* (EBV) de humanos se ha podido demostrar que la desubiquitinación producida por la proteína viral BPLF1 *in vitro* es capaz de atenuar el reclutamiento de la DNA polimerasa η en zonas del DNA dañado (Whitehurst et al., 2012). También existen trabajos que demuestran la interacción entre proteínas de geminivirus y PCNA, que es imprescindible para el reclutamiento de las TLS (Castillo et al., 2003; Luque et al., 2002).

Nuestros resultados muestran que las DNA Pol TLS PolH, PolK y Rev1 no parecen tener influencia en la acumulación de genomas víricos dentro del huésped a ninguno de los tiempos analizados lo que indicaría que no están directamente implicadas en la replicación del DNA vírico. Este resultado es esperable puesto que ninguna de estas polimerasas son procesivas por lo que es esperable que en ausencia de alguna de ellas la acumulación de virus no se vea afectada. En la misma línea están los resultados mostrados por Richter y colaboradores (Richter et al., 2016b) en los que utilizando los mutantes de TLS en *A. thaliana atrev3*, *atpolh* y *atpolk* tampoco observaron cambios en la carga viral, si bien, ellos encontraron más variabilidad entre réplicas, debido posiblemente a la inoculación por biobalística y su medición de carga viral por Southern. Nuestro estudio confirma que la agroinfección por punteadura del ápice y la posterior

cuantificación por qPCR disminuye la variabilidad intermuestral y ofrece mayor robustez al análisis que la infección por biobalística y cuantificación por Southern.

En cuanto al modo de acción de las DNA Pol TLS, existen multitud de estudios que indican la necesidad de que PCNA se monoubiquitine para que se active la ruta de reparación de daño mediante TLS (Edmunds et al., 2008; Kanao et al., 2015; Lehmann et al., 2007; Masuda et al., 2015; Zlatanou et al., 2011). Dicha ubiquitinación ocurre en la misma lisina del monómero de PCNA que la sumoilación, activando cada uno de ellos procesos diferentes dentro de la célula. Se ha demostrado que TYLCV es capaz de inhibir la sumoilación de PCNA (Castillo et al., 2004; Sánchez-Durán et al., 2011), por lo que sería lógico pensar que durante la infección por geminivirus, la ubiquitinación también se podría ver alterada, como ocurre en el caso del EBV. Este hecho afectaría a la unión de PolH y PolK a PCNA. Sin embargo, estudios recientes indican que Rev1 podría no depender de la ubiquitinación como se pensaba inicialmente, sino que Rev1 parece ser la responsable de inducir la ubiquitinación a través de su unión a Rad18, que es la proteína encargada de ubiquitinar PCNA (Hendel et al., 2011; Wang et al., 2016).

Nuestros resultados apoyan estas hipótesis iniciales, ya que en las infecciones de TYLCV de las líneas mutantes de *atpolk* y *atpolh* no se encontraron variaciones importantes en número y tipo de mutaciones, ni en la carga viral con respecto al control. Sin embargo, el análisis de las cuasiespecies de TYLCV en las dos líneas mutantes de *A. thaliana atrev1-1* y *atrev1-2* mostraron un claro descenso de las transversiones en ausencia de Rev1, más concretamente en las mutaciones asociadas a cambios por citosina, como se cabría esperar por la acción de esta DNA polimerasa. Estos cambios se mantienen tanto a tiempos cortos como largos, indicando que la actividad de esta DNA Pol TLS es continua dentro de la población. Resulta interesante observar que ni el índice de Shannon ni la diversidad nucleotídica se ven afectados por la ausencia de Rev1, lo que confirma la existencia de una multitud de mecanismos que influyen en la generación y mantenimiento de la variabilidad de la cuasiespecie y que amortigua el efecto de la ausencia de alguno de estos procesos.

Tampoco se observaron diferencias en la distribución de mutaciones ni en el número de InDels en las muestras procedentes de las líneas mutantes *atrev1-1* y *atrev1-2*. La composición de haplotipos de los espectros de mutantes obtenidos tras la

reconstrucción de la cuasiespecie de TYLCV en estas dos líneas mutantes tampoco se vio alterada, mostrando un número de haplotipos similar al control.

En cuanto al estudio de eventos de recombinación, la ausencia de diferencias entre los tipos de moléculas recombinantes halladas en las cuasiespecies generadas en plantas carentes de Rev1 y los de las plantas WT está justificada puesto que Rev1 no está implicada en ninguno de los procesos de reparación que impliquen recombinación. Es de destacar la diferencia en el número de eventos de recombinación entre cadenas de polaridad distinta con respecto a aquellos entre cadenas de la misma polaridad a 28dpi. El hecho de que las primeras sean hasta tres veces superiores en número a las segundas parece indicar un papel muy importante de la recombinación de extremos no homólogos (NHEJ). Este resultado está apoyado por la influencia de proteínas como KU80, implicadas en NHEJ, en cuya ausencia se retarda la acumulación de geminivirus (Richter and Jeske, 2015). Es por ello que no se nos escapa la necesidad de estudiar el papel que pudiesen jugar polimerasas responsables del NHEJ como Pol λ en la reparación de extremos no homólogos de DNA viral truncado. Recientemente Richter y colaboradores (Richter et al., 2016d) han actualizado el modelo de replicación de geminivirus incluyendo en el modelo la reparación de genomas truncados en la fase de replicación de la cadena complementaria (CSR) mediante recombinación homóloga, pero nuestros resultados sugieren una importancia mayor de NHEJ en la recombinación de geminivirus a tenor de los resultados hallados.

Por otra parte, el estudio de la nueva primasa-polimerasa PrimPol y su implicación sobre el ciclo replicativo de TYLCV arrojó resultados muy interesantes. Hemos mostrado que PrimPol de *A. thaliana* está implicada en el inicio de replicación de TYLCV dado que en ausencia de PrimPol la carga viral se reduce hasta un 90% a tiempos cortos de infección. Sin embargo, a partir de los 14 dpi, la carga viral ya no se ve afectada por la ausencia de esta DNA primasa-polimerasa. Esto apoya la hipótesis de la existencia de dos tipos de mecanismos de replicación diferentes en TYLCV, una replicación inicial responsable de la copia de la cadena complementaria, la CSR, dependiente de un primado hasta hoy desconocido, y una segunda replicación por círculo rodante, o RCR, independiente de primado (explicado en el apartado 5.3. de la Introducción). Según nuestros resultados, PrimPol podría ser la responsable de iniciar la CSR mediante un hipotético primado, permitiendo la síntesis de la cadena complementaria para poder expresar la proteína

REP, implicada en la RCR. El hecho de que PrimPol permita realizar un primado con dNTPs sugiere una ventaja replicativa para el virus al conferir mayor velocidad a la RCS que no necesitaría eliminar un hipotético cebador de NTPs.

Pese al descenso drástico de la carga viral a los 7 dpi en ausencia de PrimPol, la población viral termina alcanzando valores normales a los 14 dpi, lo que sugiere la existencia de un proceso CSR minoritario e independiente de PrimPol pero suficiente para poder pasar a la RCR disparando la replicación del virus. PrimPol sólo se expresa de forma abundante en células acompañantes del floema, debido a su función en mitocondrias ya que estas células presentan un mayor acúmulo de mitocondrias para favorecer el transporte activo de sustancias desde y hasta el floema. Este hecho proporciona una hipótesis acerca de la confinación de la replicación de geminivirus a esta estructura de la planta, puesto que sin expresión de PrimPol en la célula no se podría iniciar la fase de CSR.

En cuanto a la influencia de PrimPol sobre la variabilidad genética de las cuasiespecies, hemos podido observar un descenso generalizado de las mutaciones. Sin embargo, no podemos confirmar que ese descenso sea debido a la ausencia de PrimPol o al descenso del número de moléculas virales como consecuencia de dicha ausencia. Si bien a 28 dpi hemos podido observar un descenso de mutaciones T→G, en un principio no parece estar asociado al tipo de mutaciones esperables por acción de la PrimPol, ya que según estudios preliminares los cambios asociados a la actividad de PrimPol sería G→A, T→C (Guilliam et al., 2015). Sería interesante realizar experimentos *in vitro* con PrimPol y TYLCV para confirmar qué sesgo mutacional introduce esta polimerasa.

También hemos podido demostrar que es posible trabajar con replicas biológicas de cuasiespecies virales cuando se analizan muchas variables en conjunto. La nueva tecnología de NGS acoplada a nuestro programa QuasiFlow abre una nueva metodología en el estudio de las cuasiespecies. Hasta ahora los procesos aleatorios que se producían dentro de cada cuasiespecie alterando su población, como los cuellos de botella, impedían hacer comparaciones entre la mayoría de variables individuales que tradicionalmente se han analizado en el estudio de las cuasiespecies virales. La posibilidad que ofrece QuasiFlow de generar estudios de análisis de componentes principales resulta en un incremento de la reproducibilidad y robustez en el estudio de las variables además de ofrecer una visión holística de la cuasiespecie.

Por último, el hecho que tanto Rev1 como PrimPol estén implicadas en la replicación y generación de variabilidad de geminivirus, invita a pensar en nuevos procesos que podrían formar parte de la replicación y regulación de estos virus. Recientemente la investigación en el campo de la epigenética se ha interesado por los G-cuadrupeplex como mecanismo regulador de multitud de procesos como el mantenimiento de telómeros, replicación, transcripción y la traducción (Castillo-Gonzalez and Perez-Machado, 2012; Lam et al., 2013; Maizels, 2015; Sarkies et al., 2010; van Kregten and Tijsterman, 2014). Estas estructuras terciarias actuarían principalmente bloqueando el paso de la maquinaria replicativa. Recientemente se está documentando la presencia de G-cuadrupeplex en el genoma o en los transcritos de ciertos virus de DNA y RNA (Kusov et al., 2015; Métifiot et al., 2014; Murat et al., 2014; Piekna-Przybylska et al., 2014; Tlučková et al., 2013). Tanto Rev1 como PrimPol han sido descritas como las DNA polimerasas implicadas en permitir la replicación a través de los G-cuadrupeplex, pero mediante mecanismos diferentes. Rev1 es capaz de linearizar el G-cuadrupeplex permitiendo el paso de la maquinaria replicativa a través de las secuencias ricas en G (Eddy et al., 2014; van Kregten and Tijsterman, 2014; Wickramasinghe et al., 2015), mientras que PrimPol parece replicar a través de estas estructuras sin necesidad de resolverlas, pudiendo crear puntos de recombinación homóloga entre los extremos de las estructuras G-cuadrupeplexes (Schiavone et al., 2016). Los geminivirus pasan la mayor parte de su ciclo de vida en forma de ssDNA, lo que permite la formación de este tipo de estructuras terciarias. El hecho de que estas dos polimerasas estén implicadas en la resolución de estas estructuras nos llevó a hipotetizar sobre la existencia de G-cuadrupeplex en la secuencia de TYLCV. Realizamos un estudio *in silico* sobre los posibles puntos de creación de G-cuadrupeplex dentro de la secuencia de TYLCV-Mild empleando la herramienta QGRS mapper (bioinformatics.ramapo.edu/QGRS/). La **Tabla 1** indica los G-cuadrupeplex teóricos de mayor *score* hallados en la secuencia de TYLCV.

Tabla 1. Secuencias presentes en TYLCV-Mild con potencial para formar G-cuadrupeles

Posición	Tamaño	Guaninas implicadas en el G-cuadrupele	G-Score
919	28	<u>GG</u> AAACAGCCCAATGGATT <u>TTGG</u> ACAG <u>GG</u>	12
1049	24	<u>GG</u> CCCTCTGGAATGAAGGAACAGG	19
1865	15	<u>GG</u> AGGTAGGTCCTGG	18
2679	24	<u>GGGG</u> TTTCTAGGTTTTTAATTGG	12

Una observación indirecta de la existencia de estas estructuras G-cuadrupele en TYLCV se obtendría estudiando el efecto que tienen Rev1 y PrimPol sobre las secuencias que forman estas estructuras, ya que en ausencia de PrimPol deberían observarse una menor cantidad de recombinaciones homólogas en las regiones indicadas en la **Tabla 1**, mientras que en ausencia de Rev1 no debería haber cambios significativos en las recombinaciones de esa región, puesto que Rev1 es capaz de desplegar la estructura terciaria. Cuando analizamos los eventos de recombinación homóloga en las regiones acotadas por los cuatro hipotéticos G-cuadrupele pudimos observar diferencias estadísticamente significativas en ausencia de PrimPol mientras que en ausencia de Rev1 no hubo diferencias con respecto al WT para una t de Student $P=0,05$ (**Figura 1**). Esta observación, además de sugerir la existencia de G-cuadrupele en la secuencia de TYLCV, apoya la hipótesis del papel de Rev1 y PrimPol en la generación de diversidad de TYLCV-Mild.

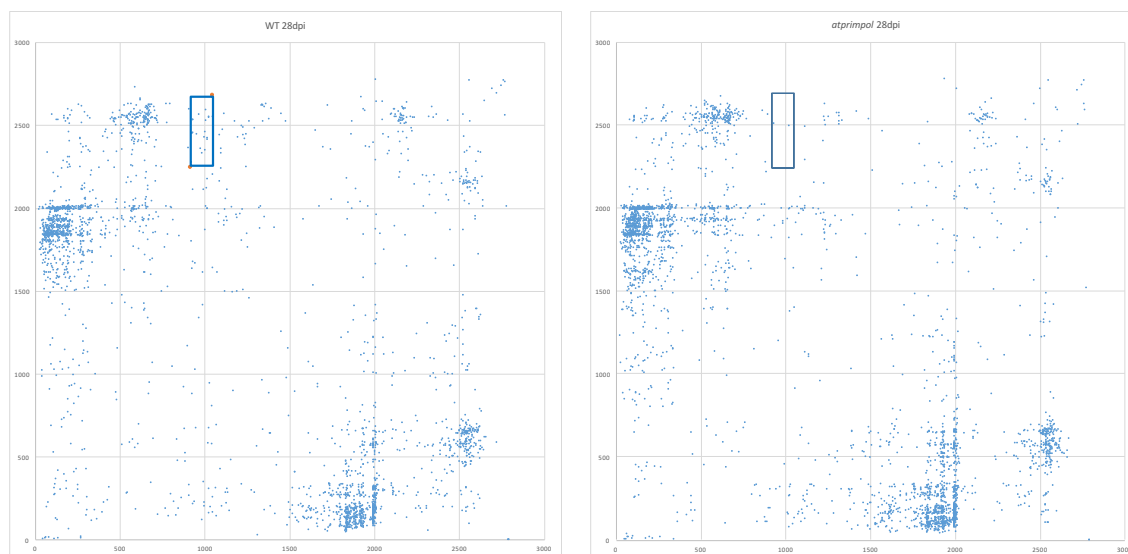


Figura 1. Figura 18. Scatter plot de las recombinaciones entre cadenas de la misma polaridad halladas en cuasiespecies de TYLCV infectando *A. thaliana* WT y mutantes *atprimpol* a 28 dpi. En ambos ejes de las gráficas se representan las posiciones nucleotídicas del genoma de TYLCV. El cuadro azul delimita el área en el que se concentran los 4 G-cuadruplex hallados mediante estudio *in silico* de la secuencia de TYLCV. El numero de puntos de las tres muestras biológicas de TYLCV WT y TYLCV crecida en la línea *atprimpol* fueron estadísticamente diferentes según T de Student para un $p < 0,05$.

CONCLUSIONES

Los resultados obtenidos en este trabajo permiten establecer las siguientes conclusiones:

1. La mutagénesis incrementada del virus de RNA TMV en el hospedador *N. tabacum* con el análogo de base 5-FU a 100 µg/ml provoca un descenso de la infectividad pero no de la carga vírica, lo que reduce su infectividad específica *in vivo* respecto al control.
2. Los espectros de mutantes de TMV mutagenizados *in vivo* no presentan cambios en su frecuencia de mutación, ni en su heterogeneidad, ni en el número medio de mutaciones por molécula, respecto a las poblaciones control. Solo hubo un cambio en la secuencia consenso de TMV en una de las poblaciones tratadas con 100 µg/ml de 5-FU a los 10 dpi, aunque desconocemos si se trata de una mutación de resistencia al 5-FU.
3. Los tipos de mutaciones predominantes en las poblaciones de TMV con reducciones en su infectividad específica son transiciones A→G y U→C, que son los cambios esperados por la incorporación del análogo al RNA vírico.
4. El tratamiento durante 10 días con 100 µg/ml de 5-FU no altera los niveles de los ribonucleótidos trifosfato intracelulares de *N. tabacum in vivo* ni se detecta la forma trifosfato del 5-FU, el FUTP. Sin embargo las células de *N. tabacum* BY-2 pueden sintetizar FUTP.
5. El 5-FU altera el equilibrio mutación-selección de las cuasiespecies de TMV así como el espacio de secuencia que estas exploran especialmente en la región RdRp, que en poblaciones control está más restringida al cambio que la región MP.
6. Los resultados apoyan el modelo de defección letal *in vivo* que explica las bases moleculares de la mutagénesis letal de TMV en el hospedador *N. tabacum*.
7. Hemos desarrollado un flujo de trabajo, QuasiFlow, que permite analizar simultáneamente y de manera sencilla, robusta y rápida la variabilidad de múltiples cuasiespecies virales a partir de datos de NGS. Esta herramienta es capaz de reconstruir haplotipos completos, detectar genomas virales recombinantes y variantes genómicas minoritarias presentes en los espectros de mutantes de coinfecciones de geminivirus.

8. La versatilidad de Quasiflow permite extender su utilización a cualquier estudio sobre variantes genómicas minoritarias en muestras de genomas con cierto grado de variabilidad como el DNA mitocondrial.

9. *AtPrimPol*, homóloga de una DNA primasa-polimerasa descrita en humanos, parece estar implicada en las fases iniciales de la infección del geminivirus TYLCV en *A. thaliana*, ya que en mutantes *AtPRIMPOL*, la carga viral desciende hasta un 90%.

10. La carga viral de TYLCV en la planta *A. thaliana* no se ve afectada en mutantes de las DNA polimerasas de translesión PolK, PolH y Rev1, lo que sugiere que estas DNA polimerasas no parecen estar implicadas en la replicación del virus.

11. La DNA polimerasa de translesión Rev1 parece tener un papel importante en la generación de la variabilidad de TYLCV, ya que en las líneas mutantes *rev1-1* y *rev1-2* de *A. thaliana* desciende el número de transversiones asociadas a la citosina en los genomas víricos. Dado que el genoma de TYLCV es de DNA monocatenario, y por tanto, más susceptible a daño, Rev1 estaría implicada en la reparación del DNA viral.

12. La DNA polimerasa de translesión Rev1 no contribuye a la variabilidad de TYLCV mediante la generación de recombinantes. Sin embargo, *AtPrimPol* sí parece estar involucrada en la recombinación de TYLCV dado que en los mutantes de esta polimerasa observamos una disminución significativa en el número de eventos de recombinación ilegítima en la región Rep de TYLCV entre cadenas de la misma polaridad. Dicha región del genoma de TYLCV contiene potenciales estructuras G-quadruplex.

13. El elevado número de recombinantes observado a 28 dpi entre cadenas complementarias del virus TYLCV sugiere la participación de polimerasas asociadas a NHEJ durante la reparación del genoma del virus.

BIBLIOGRAFÍA

BIBLIOGRAFÍA

Abel, P. P., Nelson, R. S., De, B., Hoffmann, N., Rogers, S. G., Fraley, R. T., et al. (1986). Delay of disease development in transgenic plants that express the tobacco mosaic virus coat protein gene. *Science* 232, 738–743.

Abhary, M., Patil, B. L., and Fauquet, C. M. (2007). Molecular biodiversity, taxonomy, and nomenclature of tomato yellow leaf curl-like viruses. *Tomato yellow leaf curl virus disease*. doi:10.1007/978-1-4020-4769-5_6.

Ach, R. A., Durfee, T., Miller, A. B., Taranto, P., Hanley-Bowdoin, L., Zambryski, P. C., et al. (1997). RRB1 and RRB2 encode maize retinoblastoma-related proteins that interact with a plant D-type cyclin and geminivirus replication protein. *Mol Cell Biol.* 17 17.

Acosta-Leal, R., Duffy, S., Xiong, Z., Hammond, R. W., and Elena, S. F. (2011). Advances in Plant Virus Evolution: Translating Evolutionary Insights into Better Disease Management. <http://dx.doi.org/10.1094/PHYTO-01-11-0017> 101, 1136–1148. doi:10.1094/PHYTO-01-11-0017.

Ahlquist, P., Noueir, A. O., Lee, W.-M., Kushner, D. B., and Dye, B. T. (2003). Host factors in positive-strand RNA virus genome replication. *Journal of Virology* 77, 8181–8186. doi:10.1128/JVI.77.15.8181-8186.2003.

Airaksinen, A., Pariente, N., Menéndez-Arias, L., and Domingo, E. (2003). Curing of foot-and-mouth disease virus from persistently infected cells by ribavirin involves enhanced mutagenesis. *Virology* 311, 339–349. doi:10.1016/S0042-6822(03)00144-2.

Allison, R. F., Janda, M., and Ahlquist, P. (1989). Sequence of cowpea chlorotic mottle virus RNAs 2 and 3 and evidence of a recombination event during bromovirus evolution. *Virology* 172, 321–330.

Amoroso, A., Concia, L., Maggio, C., Raynaud, C., Bergounioux, C., Crespan, E., et al. (2011). Oxidative DNA Damage Bypass in *Arabidopsis thaliana* Requires DNA Polymerase and Proliferating Cell Nuclear Antigen 2. *The Plant Cell* 23, 806–822. doi:10.1105/tpc.110.081455.

An, G. (1987). Binary Ti vectors for plant transformation and promoter analysis. *Methods in Enzymology. Recombinant DNA*. 153, 292–305.

Anderson, H. J., Vonarx, E. J., Pastushok, L., Nakagawa, M., Katafuchi, A., Gruz, P., et al. (2008). *Arabidopsis thaliana* Y-family DNA polymerase eta catalyses translesion synthesis and interacts functionally with PCNA2. *The Plant Journal* 55, 895–908. doi:10.1111/j.1365-313X.2008.03562.x.

Anderson, K. B., Guest, J. L., and Rimland, D. (2004). Hepatitis C virus coinfection increases mortality in HIV-infected patients in the highly active antiretroviral therapy era: data from the HIV Atlanta VA Cohort Study. *Clin. Infect. Dis.* 39, 1507–1513. doi:10.1086/425360.

Andino, R., and Domingo, E. (2015). Viral quasispecies. *Virology* 479-480C, 46–51. doi:10.1016/j.virol.2015.03.022.

Aoki, S., and Takebe, I. (1969). Infection of tobacco mesophyll protoplasts by tobacco mosaic virus ribonucleic acid. *Virology* 39, 439–448. doi:10.1016/0042-6822(69)90092-0.

Archer, J., Weber, J., Henry, K., Winner, D., Gibson, R., Lee, L., et al. (2012). Use of four next-generation sequencing platforms to determine HIV-1 coreceptor tropism. *PLoS ONE* 7, e49602. doi:10.1371/journal.pone.0049602.

Argüello-Astorga, G., Herrera-Estrella, L., and Rivera-Bustamante, R. (1994). Experimental and theoretical definition of geminivirus origin of replication. *Plant Mol Biol* 26, 553–556.

Arias, A., Isabel de Ávila, A., Sanz-Ramos, M., Agudo, R., Escarmís, C., and Domingo, E. (2013). Molecular dissection of a viral quasispecies under mutagenic treatment: positive correlation between fitness loss and mutational load. *J. Gen. Virol.* 94, 817–830. doi:10.1099/vir.0.049171-0.

Arias, A., Thorne, L., and Goodfellow, I. (2014). Favipiravir elicits antiviral mutagenesis during virus replication in vivo . *eLife* 3. doi:10.7554/eLife.03679.

Armenia, D., Vandenbroucke, I., Fabeni, L., Van Marck, H., Cento, V., D'Arrigo, R., et al. (2012). Study of genotypic and phenotypic HIV-1 dynamics of integrase mutations during raltegravir treatment: a refined analysis by ultra-deep 454 pyrosequencing. *J. Infect. Dis.* 205, 557–567. doi:10.1093/infdis/jir821.

Asahina, Y., Izumi, N., Enomoto, N., Uchihara, M., Kurosaki, M., Onuki, Y., et al. (2005). Mutagenic effects of ribavirin and response to interferon/ribavirin combination therapy in chronic hepatitis C. *J. Hepatol.* 43, 623–629. doi:10.1016/j.jhep.2005.05.032.

Atkins, D., Hull, R., Wells, B., Roberts, K., Moore, P., and Beachy, R. N. (1991). The tobacco mosaic virus 30K movement protein in transgenic tobacco plants is localized to plasmodesmata. *Journal of General Virology* 72 (Pt 1), 209–211. doi:10.1099/0022-1317-72-1-209.

Avkin, S., and Livneh, Z. (2002). Efficiency, specificity and DNA polymerase-dependence of translesion replication across the oxidative DNA lesion 8-oxoguanine in human cells. *Mutat. Res.* 510, 81–90.

Azzam, O., and Frazer, J. (1994). de la Rosa, D., Beaver, J. S., Ahlquist, P., Maxwell, D. P. 1994. *Whitefly transmission and efficient ssDNA accumulation of bean golden mosaic geminivirus require functional coat protein.* *Virology*.

Baer, R., Bankier, A. T., Biggin, M. D., Deininger, P. L., Farrell, P. J., Gibson, T. J., et al. (1984). DNA sequence and expression of the B95-8 Epstein-Barr virus genome. *Nature* 310, 207–211.

Bagewadi, B., Chen, S., Lal, S. K., Choudhury, N. R., and Mukherjee, S. K. (2004). PCNA interacts with Indian mung bean yellow mosaic virus rep and downregulates Rep activity. *Journal of Virology* 78, 11890–11903. doi:10.1128/JVI.78.21.11890-11903.2004.

Bar-Ziv, A., Levy, Y., Citovsky, V., and Gafni, Y. (2015). The Tomato yellow leaf curl virus (TYLCV) V2 protein inhibits enzymatic activity of the host papain-like cysteine protease CYP1. *Biochem Biophys Res Commun.* 460 460.

Bar-Ziv, A., Levy, Y., Hak, H., Mett, A., Belausov, E., Citovsky, V., et al. (2012). The tomato yellow leaf curl virus (TYLCV) V2 protein interacts with the host papain-like cysteine protease CYP1. *Plant Signal Behav.*

Baranowski, E., Ruiz-Jarabo, C. M., Pariente, N., Verdaguer, N., and Domingo, E. (2003). Evolution of cell recognition by viruses: a source of biological novelty with medical implications. *Adv. Virus Res.* 62, 19–111.

Barba, M., Czosnek, H., and Hadidi, A. (2014). Historical Perspective, Development and Applications of Next-Generation Sequencing in Plant Virology. *Viruses* 6, 106–136. doi:10.3390/v6010106.

Barta, A., Sum, I., and Föglein, F. J. (1981). 2-Thiouracil Does Not Inhibit TMV Replication in Tobacco Protoplasts. *J. Gen. Virol.* 56, 219–222. doi:10.1099/0022-1317-56-1-219.

Barzon, L., Lavezzo, E., Costanzi, G., Franchin, E., Toppo, S., and Palù, G. (2013). Next-generation sequencing technologies in diagnostic virology. *J. Clin. Virol.* 58, 346–350. doi:10.1016/j.jcv.2013.03.003.

Bândeă, C. I. (1983). A new theory on the origin and the nature of viruses. *Journal of Theoretical Biology* 105, 591–602. doi:10.1016/0022-5193(83)90221-7.

Beier, H., Barciszewska, M., Krupp, G., Mitnacht, R., and Gross, H. J. (1984). UAG readthrough during TMV RNA translation: isolation and sequence of two tRNAs with suppressor activity from tobacco plants. *EMBO J.* 3, 351–356.

Beijerinck, M. W. (1898). *Over een contagium vivum fluidum also oorzaak van de Vleckziek te der tabaksbladen*. Verslag Koninkl. Akad. van Wetensch. te Amsterdam.

Bennett, S. (2004). Solexa Ltd. *Pharmacogenomics* 5, 433–438. doi:10.1517/14622416.5.4.433.

Bernal, J. D., and Fankuchen, I. (1941). X-ray and crystallographic studies of plant virus preparations : i. Introduction and preparation of specimens ii. modes of aggregation of the virus particles. *J. Gen. Physiol.* 25, 111–146.

Bianchi, J., Rudd, S. G., Jozwiakowski, S. K., Bailey, L. J., Soura, V., Taylor, E., et al. (2013). PrimPol bypasses UV photoproducts during eukaryotic chromosomal DNA replication. *Mol. Cell* 52, 566–573. doi:10.1016/j.molcel.2013.10.035.

Biebricher, C. K., and Eigen, M. (2006). What is a quasispecies? *Curr. Top. Microbiol. Immunol.* 299, 1–31.

Biertümpfel, C., Zhao, Y., Kondo, Y., Ramón-Maiques, S., Gregory, M., Lee, J. Y., et al. (2010a). Structure and mechanism of human DNA polymerase ϵ . *Nature* 465, 1044–1048. doi:10.1038/nature09196.

Biertümpfel, C., Zhao, Y., Kondo, Y., Ramón-Maiques, S., Gregory, M., Lee, J. Y., et al. (2010b). Structure and mechanism of human DNA polymerase ϵ . *Nature* 465, 1044–1048. doi:10.1038/nature09196.

Bisaro, D. M. (2006). Silencing suppression by geminivirus proteins. *Virology* 344 344.

Bordería, A. V., Lorenzo-Redondo, R., Pernas, M., Casado, C., Alvaro, T., Domingo, E., et al. (2010). Initial fitness recovery of HIV-1 is associated with quasispecies heterogeneity and can occur without modifications in the consensus sequence. *PLoS ONE* 5, e10319. doi:10.1371/journal.pone.0010319.

Boudsocq, F., Kokoska, R. J., Plosky, B. S., Vaisman, A., Ling, H., Kunkel, T. A., et al. (2004). Investigating the role of the little finger domain of Y-family DNA polymerases in low fidelity synthesis and translesion replication. *Journal of Biological Chemistry* 279, 32932–32940. doi:10.1074/jbc.M405249200.

Bousalem, M., Douzery, E. J., and Fargette, D. (2000). High genetic diversity, distant phylogenetic relationships and intraspecies recombination events among natural populations of Yam mosaic virus: a contribution to understanding potyvirus evolution. *Journal of General Virology* 81, 243–255. doi:10.1099/0022-1317-81-1-243.

Bowers, J., Mitchell, J., Beer, E., Buzby, P. R., Causey, M., Efcavitch, J. W., et al. (2009). Virtual terminator nucleotides for next-generation DNA sequencing. *Nat. Methods* 6, 593–595. doi:10.1038/nmeth.1354.

Brandizzi, F., and Wasteneys, G. O. (2013). Cytoskeleton-dependent endomembrane organization in plant cells: an emerging role for microtubules. *The Plant Journal* 75, 339–349. doi:10.1111/tpj.12227.

Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D. H., Johnson, D., et al. (2000). Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat. Biotechnol.* 18, 630–634. doi:10.1038/76469.

Briddon, R. W., Bedford, I. D., Tsai, J. H., and Markham, P. G. (1996). Analysis of the nucleotide sequence of the treehopper-transmitted geminivirus, tomato pseudo-curly top virus, suggests a recombinant origin. *Virology* 219, 387–394. doi:10.1006/viro.1996.0264.

Brochot, E., Duverlie, G., Castelain, S., Morel, V., Wychowski, C., Dubuisson, J., et al. (2007). Effect of ribavirin on the hepatitis C virus (JFH-1) and its correlation with interferon sensitivity. *Antivir. Ther. (Lond.)* 12, 805–813.

Brown, J. K., Zerbini, F. M., Navas-Castillo, J., Moriones, E., Ramos-Sobrinho, R., Silva, J. C. F., et al. (2015). Revision of Begomovirus taxonomy based on pairwise sequence comparisons. *Arch Virol* 160, 1593–1619. doi:10.1007/s00705-015-2398-y.

Brumme, C. J., Huber, K. D., Dong, W., Poon, A. F. Y., Harrigan, P. R., and Sluis-Cremer, N. (2013). Replication fitness of multiple nonnucleoside reverse transcriptase-resistant HIV-1 variants in the presence of etravirine measured by 454 deep sequencing. *Journal of Virology* 87, 8805–8807. doi:10.1128/JVI.00335-13.

Buchmann, R. C., Asad, S., Wolf, J. N., Mohannath, G., and Bisaro, D. M. (2009). Geminivirus AL2 and L2 proteins suppress transcriptional gene silencing and cause genome-wide reductions in cytosine methylation. *J. Virol.*

Bull, R. A., Luciani, F., McElroy, K., Gaudieri, S., Pham, S. T., Chopra, A., et al. (2011). Sequential bottlenecks drive viral evolution in early acute hepatitis C virus infection. *PLoS Pathog* 7, e1002243. doi:10.1371/journal.ppat.1002243.

Bull, S. E., Briddon, R. W., Sserubombwe, W. S., Ngugi, K., Markham, P. G., and Stanley, J. (2006). Genetic diversity and phylogeography of cassava mosaic viruses in Kenya. *Journal of General Virology* 87, 3053–3065. doi:10.1099/vir.0.82013-0.

Burgers, P. M., Koonin, E. V., Bruford, E., Blanco, L., Burtis, K. C., Christman, M. F., et al. (2001). Eukaryotic DNA polymerases: proposal for a revised nomenclature. *Journal of Biological Chemistry* 276, 43487–43490. doi:10.1074/jbc.R100056200.

Bustin, S. A., Benes, V., Garson, J. A., Hellemans, J., Huggett, J., Kubista, M., et al. (2009). The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. *Clinical Chemistry* 55, 611–622. doi:10.1373/clinchem.2008.112797.

Cañizares, M. C., Rosas Díaz, T., Rodríguez Negrete, E., Hogenhout, S. A., Bedford, I. D., Bejarano, E. R., et al. (2015). Arabidopsis thaliana, an experimental host for tomato yellow leaf curl disease-associated begomoviruses by agroinoculation and whitefly transmission. *Plant Pathology* 64, 265–271. doi:10.1111/ppa.12270.

Carr, J. P., Marsh, L. E., Lomonossoff, G. P., Sekiya, M. E., and Zaitlin, M. (1992). Resistance to tobacco mosaic virus induced by the 54-kDa gene sequence requires expression of the 54-kDa protein. *MPMI* 5, 397–404.

Caspar, D. L. (1956). Structure of bushy stunt virus. *Nature* 177, 475–476.

Castillo, A. G., Collinet, D., Deret, S., Kashoggi, A., and Bejarano, E. R. (2003). Dual interaction of plant PCNA with geminivirus replication accessory protein (Ren) and viral replication protein (Rep). *Virology*. 312 312.

Castillo, A. G., Kong, L. J., Hanley-Bowdoin, L., and Bejarano, E. R. (2004). Interaction between a geminivirus replication protein and the plant sumoylation system. *Journal of Virology* 78, 2758–2769. doi:10.1128/JVI.78.6.2758-2769.2004.

Castillo-Gonzalez, D., and Perez-Machado, G. (2012). Computational tools in the discovery of new G-quadruplex ligands with potential anticancer activity. *Current topics in ...*

Chao, L., Rang, C. U., and Wong, L. E. (2002). Distribution of spontaneous mutants and inferences about the replication mode of the RNA bacteriophage phi6. *Journal of Virology* 76, 3276–3281. doi:10.1128/JVI.76.7.3276-3281.2002.

Chapman, S. N. (1998). Tobamovirus isolation and RNA extraction. *Methods Mol. Biol.* 81, 123–129. doi:10.1385/0-89603-385-6:123.

Chelico, L., Pham, P., Calabrese, P., and Goodman, M. F. (2006). APOBEC3G DNA deaminase acts processively 3' → 5' on single-stranded DNA. *Nat. Struct. Mol. Biol.* 13, 392–399. doi:10.1038/nsmb1086.

Chenault, K. D., and Melcher, U. (1994). Phylogenetic relationships reveal recombination among isolates of cauliflower mosaic virus. *J. Mol. Evol.* 39, 496–505.

Chevaliez, S., Brillet, R., Lázaro, E., Hézode, C., and Pawlotsky, J.-M. (2007). Analysis of ribavirin mutagenicity in human hepatitis C virus infection. *Journal of Virology* 81, 7732–7741. doi:10.1128/JVI.00382-07.

Chung, H. J., Lee, J.-W., Kim, Y. S., and Lee, J. I. (2013). Prediction of sustained virologic response based on week 4 and week 12 response in hepatitis C virus genotype 1 patients treated with peginterferon and ribavirin: assessment in a favorable IL28B allele-prevalent area. *Intervirology* 56, 178–183. doi:10.1159/000345539.

Clérot, D., and Bernardi, F. (2006). DNA helicase activity is associated with the replication initiator protein rep of tomato yellow leaf curl geminivirus. *Journal of Virology* 80, 11322–11330. doi:10.1128/JVI.00924-06.

Clouser, C. L., Holtz, C. M., Mullett, M., Crankshaw, D. L., Briggs, J. E., Chauhan, J., et al. (2011). Analysis of the ex vivo and in vivo antiretroviral activity of gemcitabine. *PLoS ONE* 6, e15840. doi:10.1371/journal.pone.0015840.

Clouser, C. L., Patterson, S. E., and Mansky, L. M. (2010). Exploiting drug repositioning for discovery of a novel HIV combination therapy. *Journal of Virology* 84, 9301–9309. doi:10.1128/JVI.01006-10.

Contreras, A. M., Hiasa, Y., He, W., Terella, A., Schmidt, E. V., and Chung, R. T. (2002). Viral RNA mutations are region specific and increased by ribavirin in a full-length hepatitis C virus replication system. *Journal of Virology* 76, 8505–8517. doi:10.1128/JVI.76.17.8505-8517.2002.

Corchado, E., Corchado, J. M., and Abraham, A. eds. (2007). *Innovations in Hybrid Intelligent Systems*. Berlin, Heidelberg: Springer Berlin Heidelberg doi:10.1007/978-3-540-74972-1.

Crotty, S., Maag, D., Arnold, J. J., Zhong, W., Lau, J. Y., Hong, Z., et al. (2000). The

broad-spectrum antiviral ribonucleoside ribavirin is an RNA virus mutagen. *Nat. Med.* 6, 1375–1379. doi:10.1038/82191.

Crowder, S., and Kirkegaard, K. (2005). Trans-dominant inhibition of RNA viral replication can slow growth of drug-resistant viruses. *Nat. Genet.* 37, 701–709. doi:10.1038/ng1583.

Dapp, M. J., Clouser, C. L., Patterson, S., and Mansky, L. M. (2009). 5-Azacytidine can induce lethal mutagenesis in human immunodeficiency virus type 1. *Journal of Virology* 83, 11950–11958. doi:10.1128/JVI.01406-09.

Davern, C. I., and Bonner, J. (1958). The influence of 5-fluorouracil on tobacco-mosaic virus production in tobacco-leaf discs. *Biochim. Biophys. Acta* 29, 205–206.

Dawson, W. O., and Boyd, C. (1987). Modifications of nucleic acid precursors that inhibit plant virus multiplication. *Phytopathology* 77, 477–480.

Dawson, W. O., and Grantham, G. L. (1983). Effect of 2-thiouracil on RNA and protein synthesis in synchronous and asynchronous infections of tobacco mosaic virus. *Intervirology* 19, 155–161.

Dawson, W. O., and Lehto, K. M. (1990). Regulation of tobamovirus gene expression. *Adv. Virus Res.* 38, 307–342.

Dawson, W. O., and Lozoya-Saldana, H. (1984). Examination of the mode of action of ribavirin against tobacco mosaic virus. *Intervirology* 22, 77–84.

Dawson, W. O., and Painter, P. R. (1978). Estimation of the cell-to-cell spread of tobacco mosaic virus infection in mechanically inoculated leaves. *Intervirology* 9, 310–315.

Dawson, W. O., German, T. L., and Schlegel, D. E. (1976). Homogenization-resistant and -susceptible components of tobacco mosaic virus replicative form RNA. *Journal of General Virology* 32, 205–215. doi:10.1099/0022-1317-32-2-205.

Day, M. J. D., Ashurst, J. A., Mathias, S. F., Watts, J. W., Wilson, T. M. A., and Dixon, R. A. (1993). Plant viral leaders influence expression of a reporter gene in tobacco. *Plant Mol Biol* 23, 97–109. doi:10.1007/BF00021423.

Delatte, H., Holota, H., Moury, B., Reynaud, B., Lett, J.-M., and Peterschmitt, M. (2007). Evidence for a founder effect after introduction of Tomato yellow leaf curl virus-mild in an insular environment. *J. Mol. Evol.* 65, 112–118. doi:10.1007/s00239-007-0005-x.

Deom, C. M., Oliver, M. J., and Beachy, R. N. (1987). The 30-kilodalton gene product of tobacco mosaic virus potentiates virus movement. *Science* 237, 389–394. doi:10.1126/science.237.4813.389.

Dietz, J., Schelhorn, S.-E., Fitting, D., Mihm, U., Susser, S., Welker, M.-W., et al. (2013). Deep sequencing reveals mutagenic effects of ribavirin during monotherapy of

hepatitis C virus genotype 1-infected patients. *Journal of Virology* 87, 6172–6181. doi:10.1128/JVI.02778-12.

Domingo, E., Dávila, M., and Ortín, J. (1980). Nucleotide sequence heterogeneity of the RNA from a natural population of foot-and-mouth-disease virus. *Gene* 11, 333–346.

Domingo, E., Martin, V., Perales, C., Grande-Perez, A., García-Arriaza, J., and Arias, A. (2006). Viruses as quasispecies: biological implications. *Curr. Top. Microbiol. Immunol.* 299, 51–82.

Domingo, E., Sabo, D., Taniguchi, T., and Weissmann, C. (1978). Nucleotide sequence heterogeneity of an RNA phage population. *Cell* 13, 735–744.

Domingo, E., Sheldon, J., and Perales, C. (2012). Viral quasispecies evolution. *Microbiol. Mol. Biol. Rev.* 76, 159–216. doi:10.1128/MMBR.05023-11.

Dong, X., van Wezel, R., Stanley, J., and Hong, Y. (2003). Functional characterization of the nuclear localization signal for a suppressor of posttranscriptional gene silencing. *Journal of Virology* 77, 7026–7033. doi:10.1128/JVI.77.12.7026-7033.2003.

Drake, J. W. (1991). A constant rate of spontaneous mutation in DNA-based microbes. *Proc Natl Acad Sci USA* 88, 7160–7164.

Drake, J. W. (1993). Rates of spontaneous mutation among RNA viruses. *Proc Natl Acad Sci USA* 90, 4171–4175.

Drake, J. W., and Hwang, C. B. C. (2005). On the mutation rate of herpes simplex virus type 1. *Genetics* 170, 969–970. doi:10.1534/genetics.104.040410.

Drake, J. W., Charlesworth, B., Charlesworth, D., and Crow, J. F. (1998). Rates of spontaneous mutation. *Genetics* 148, 1667–1686.

Drummond, A. J., Pybus, O. G., Rambaut, A., Forsberg, R., and Rodrigo, A. G. (2003). Measurably evolving populations. *Trends in Ecology & Evolution* 18, 481–488. doi:10.1016/S0169-5347(03)00216-7.

Duffy, S., and Holmes, E. C. (2008). Phylogenetic evidence for rapid rates of molecular evolution in the single-stranded DNA begomovirus tomato yellow leaf curl virus. *Journal of Virology* 82, 957–965. doi:10.1128/JVI.01929-07.

Duffy, S., Shackelton, L. A., and Holmes, E. C. (2008). Rates of evolutionary change in viruses: patterns and determinants. *Nat. Rev. Genet.* 9, 267–276. doi:10.1038/nrg2323.

Dunning, A. M., Talmud, P., and Humphries, S. E. (1988). Errors in the polymerase chain reaction. *Nucleic Acids Research* 16, 10393.

Eddy, S., Ketkar, A., Zafar, M. K., Maddukuri, L., Choi, J. Y., and Eoff, R. L. (2014). Human Rev1 polymerase disrupts G-quadruplex DNA. *Nucleic Acids Research* 42, 3272–3285. doi:10.1093/nar/gkt1314.

- Edmunds, C. E., Simpson, L. J., and Sale, J. E. (2008). PCNA Ubiquitination and REV1 Define Temporally Distinct Mechanisms for Controlling Translesion Synthesis in the Avian Cell Line DT40. *Mol. Cell* 30, 519–529. doi:10.1016/j.molcel.2008.03.024.
- Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., et al. (2009). Real-time DNA sequencing from single polymerase molecules. *Science* 323, 133–138. doi:10.1126/science.1162986.
- Eigen, M. (1971). Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* 58, 465–523.
- Eigen, M. (2002). Error catastrophe and antiviral strategy. *Proc Natl Acad Sci USA* 99, 13374–13376. doi:10.1073/pnas.212514799.
- Eigen, M., and Schuster, P. (1979). *The Hyper cycle, Berlin (W.), 1979 GILLESPIE, DT: A Pedestrian Approach to Transitions and Fluctuations in Simple Nonequilibrium Chemical Systems*. Physica A.
- Escarmís, C., Dávila, M., Charpentier, N., Bracho, A., Moya, A., and Domingo, E. (1996). Genetic Lesions Associated with Muller's Ratchet in an RNA Virus. *Journal of Molecular Biology* 264, 255–267. doi:10.1006/jmbi.1996.0639.
- Escobar-Gutiérrez, A., Vazquez-Pichardo, M., Cruz-Rivera, M., Rivera-Osorio, P., Carpio-Pedroza, J. C., Ruíz-Pacheco, J. A., et al. (2012). Identification of hepatitis C virus transmission using a next-generation sequencing approach. *Journal of Clinical Microbiology* 50, 1461–1463. doi:10.1128/JCM.00005-12.
- Ewing, B., Hillier, L., Wendl, M. C., and Green, P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* 8, 175–185.
- Excoffier, L., Laval, G., and Schneider, S. (2005). Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol. Bioinform. Online* 1, 47–50.
- Excoffier, L., Smouse, P. E., and Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131, 479–491.
- F C Bawden, N. W. P. (1936). Experiments on the Chemical Behaviour of Potato Virus "X." *British Journal of Experimental Pathology* 17, 64.
- Fabre, F., Montarry, J., Coville, J., Senoussi, R., Simon, V., and Moury, B. (2012). Modelling the Evolutionary Dynamics of Viruses within Their Hosts: A Case Study Using High-Throughput Sequencing. *PLoS Pathog* 8, e1002654. doi:10.1371/journal.ppat.1002654.
- Falgueras, J., Lara, A. J., Fernández-Pozo, N., Cantón, F. R., Pérez-Trabado, G., and Claros, M. G. (2010). SeqTrim: a high-throughput pipeline for pre-processing any type of sequence read. *BMC Bioinformatics* 11, 1. doi:10.1186/1471-2105-11-38.

Fauquet, C. M., Sawyer, S., Idris, A. M., and Brown, J. K. (2005). Sequence analysis and classification of apparent recombinant begomoviruses infecting tomato in the Nile and Mediterranean basins. *Phytopathology* 95, 549–555. doi:10.1094/PHYTO-95-0549.

Feng, D. F., Johnson, M. S., and Doolittle, R. F. (1984). Aligning amino acid sequences: comparison of commonly used methods. *J. Mol. Evol.* 21, 112–125.

Fernández-Cuartero, B., Burguán, J., Aranda, M. A., Salánki, K., Moriones, E., and García-Arenal, F. (1994). Increase in the relative fitness of a plant virus RNA associated with its recombinant nature. *Virology* 203, 373–377. doi:10.1006/viro.1994.1496.

Fisher, R., van Zyl, G. U., Travers, S. A. A., Kosakovsky Pond, S. L., Engelbrecht, S., Murrell, B., et al. (2012). Deep sequencing reveals minor protease resistance mutations in patients failing a protease inhibitor regimen. *Journal of Virology* 86, 6231–6237. doi:10.1128/JVI.06541-11.

Flint, M., Mullen, S., Deatly, A. M., Chen, W., Miller, L. Z., Ralston, R., et al. (2009). Selection and characterization of hepatitis C virus replicons dually resistant to the polymerase and protease inhibitors HCV-796 and boceprevir (SCH 503034). *Antimicrob. Agents Chemother.* 53, 401–411. doi:10.1128/AAC.01081-08.

Forterre, P. (2010). Defining life: the virus viewpoint. *Orig Life Evol Biosph* 40, 151–160. doi:10.1007/s11084-010-9194-1.

Fraenkel-Conrat, H., and Williams, R. C. (1955). Reconstitution of active tobacco mosaic virus from its inactive protein and nucleic acid components. *Proc Natl Acad Sci USA* 41, 690–698.

Fraile, A., Escriu, F., Aranda, M. A., Malpica, J. M., Gibbs, A. J., and García-Arenal, F. (1997). A century of tobamovirus evolution in an Australian population of *Nicotiana glauca*. *Journal of Virology* 71, 8316–8320.

FRANKLIN, R. E., and KLUG, A. (1956). The nature of the helical groove on the tobacco mosaic virus particle; x-ray diffraction studies. *Biochim. Biophys. Acta* 19, 403–416.

French, R., and Stenger, D. C. (2003). Evolution of Wheat streak mosaic virus: dynamics of population growth within plants may explain limited variation. *Annu Rev Phytopathol* 41, 199–214. doi:10.1146/annurev.phyto.41.052002.095559.

French, R., and Stenger, D. C. (2005). Population structure within lineages of Wheat streak mosaic virus derived from a common founding event exhibits stochastic variation inconsistent with the deterministic quasi-species model. *Virology* 343, 179–189. doi:10.1016/j.virol.2005.08.037.

Fritsch, E. S., Chabbert, C. D., Klaus, B., and Steinmetz, L. M. (2014). A genome-wide map of mitochondrial DNA recombination in yeast. *Genetics* 198, 755–771. doi:10.1534/genetics.114.166637.

Froissart, R., Roze, D., Uzest, M., Galibert, L., Blanc, S., and Michalakakis, Y. (2005). Recombination every day: abundant recombination in a virus during a single multi-cellular host infection. *PLOS Biol* 3, e89. doi:10.1371/journal.pbio.0030089.

Fujiki, M., Kawakami, S., Kim, R. W., and Beachy, R. N. (2006). Domains of tobacco mosaic virus movement protein essential for its membrane association. *Journal of General Virology* 87, 2699–2707. doi:10.1099/vir.0.81936-0.

Fukunaga, R., and Doudna, J. A. (2009). dsRNA with 5' overhangs contributes to endogenous and antiviral RNA silencing pathways in plants. *EMBO J.* 28, 545–555. doi:10.1038/emboj.2009.2.

Gafni, Y. (2003). Tomato yellow leaf curl virus, the intracellular dynamics of a plant DNA virus. *Molecular Plant Pathology* 4, 9–15. doi:10.1046/j.1364-3703.2003.00147.x.

Gallie, D. R., Sleat, D. E., Watts, J. W., Turner, P. C., and Wilson, T. M. A. (1987). The 5'-leader sequence of tobacco mosaic virus RNA enhances the expression of foreign gene transcripts in vitro and in vivo. *Nucleic Acids Research* 15, 3257–3273. doi:10.1093/nar/15.8.3257.

Garcia-Diaz, M., and Bebenek, K. (2007). Multiple functions of DNA polymerases. *CRC Crit Rev Plant Sci* 26, 105–122. doi:10.1080/07352680701252817.

García-Andrés, S., Accotto, G. P., Navas-Castillo, J., and Moriones, E. (2007). Founder effect, plant host, and recombination shape the emergent population of begomoviruses that cause the tomato yellow leaf curl disease in the Mediterranean basin. *Virology* 359, 302–312. doi:10.1016/j.virol.2006.09.030.

García-Andrés, S., Monci, F., Navas-Castillo, J., and Moriones, E. (2006). Begomovirus genetic diversity in the native plant reservoir *Solanum nigrum*: Evidence for the presence of a new virus species of recombinant nature. *Virology* 350, 433–442. doi:10.1016/j.virol.2006.02.028.

García-Andrés, S., Tomás, D. M., Navas-Castillo, J., and Moriones, E. (2009). Resistance-driven selection of begomoviruses associated with the tomato yellow leaf curl disease. *Virus Research* 146, 66–72. doi:10.1016/j.virusres.2009.08.012.

García-Arenal, F., and McDonald, B. A. (2007). An Analysis of the Durability of Resistance to Plant Viruses. <http://dx.doi.org/10.1094/PHYTO.2003.93.8.941> 93, 941–952. doi:10.1094/PHYTO.2003.93.8.941.

García-Gómez, S., Reyes, A., Martínez-Jiménez, M. I., Chocrón, E. S., Mourón, S., Terrados, G., et al. (2013). PrimPol, an archaic primase/polymerase operating in human cells. *Mol. Cell* 52, 541–553. doi:10.1016/j.molcel.2013.09.025.

García-Ortiz, M. V., Ariza, R. R., Hoffman, P. D., Hays, J. B., and Roldán-Arjona, T. (2004). *Arabidopsis thaliana* AtPOLK encodes a DinB-like DNA polymerase that extends mispaired primer termini and is highly expressed in a variety of tissues. *The Plant Journal* 39, 84–97. doi:10.1111/j.1365-3113.2004.02112.x.

Ge, L., Zhang, J., Zhou, X., and Li, H. (2007). Genetic structure and population variability of tomato yellow leaf curl China virus. *Journal of Virology* 81, 5902–5907. doi:10.1128/JVI.02431-06.

Ge, P., and Zhou, Z. H. (2011). Hydrogen-bonding networks and RNA bases revealed by cryo electron microscopy suggest a triggering mechanism for calcium switches. *Proc. Natl. Acad. Sci. U.S.A.* 108, 9637–9642. doi:10.1073/pnas.1018104108.

Gerlach, V. L., Aravind, L., Gotway, G., Schultz, R. A., Koonin, E. V., and Friedberg, E. C. (1999). Human and mouse homologs of Escherichia coli DinB (DNA polymerase IV), members of the UmuC/DinB superfamily. *Proc Natl Acad Sci USA* 96, 11922–11927.

Gibbs, M. J., and Cooper, J. I. (1995). A recombinational event in the history of luteoviruses probably induced by base-pairing between the genomes of two distinct viruses. *Virology* 206, 1129–1132. doi:10.1006/viro.1995.1037.

Gibbs, P. E., Wang, X. D., Li, Z., McManus, T. P., McGregor, W. G., Lawrence, C. W., et al. (2000). The function of the human homolog of Saccharomyces cerevisiae REV1 is required for mutagenesis induced by UV light. *Proc Natl Acad Sci USA* 97, 4186–4191.

Gibson, R. M., Schmotzer, C. L., and Quiñones-Mateu, M. E. (2014). Next-Generation Sequencing to Help Monitor Patients Infected with HIV: Ready for Clinical Use? *Curr Infect Dis Rep* 16, 401–9. doi:10.1007/s11908-014-0401-5.

Gierer, A., and Mundry, K. W. (1958). Production of mutants of tobacco mosaic virus by chemical alteration of its ribonucleic acid in vitro. *Nature* 182, 1457–1458.

Gierer, A., and Schramm, G. (1956). Infectivity of ribonucleic acid from tobacco mosaic virus. *Nature* 177, 702–703.

Gladfelter, H. J., Eagle, P. A., Fontes, E. P., Batts, L., and Hanley-Bowdoin, L. (1997). Two domains of the AL1 protein mediate geminivirus origin recognition. *Virology* 239, 186–197. doi:10.1006/viro.1997.8869.

Gleason, M. K., and Fraenkel-Conrat, H. (1976). Biological consequences of incorporation of 5-fluorocytidine in the RNA of 5-fluorouracil-treated eukaryotic cells. *Proc Natl Acad Sci USA* 73, 1528–1531.

González- López, C., Arias, A., Pariente, N., Gómez-Mariano, G., and Domingo, E. (2004). Preextinction viral RNA can interfere with infectivity. *Journal of Virology* 78, 3319–3324. doi:10.1128/JVI.78.7.3319-3324.2004.

Görzer, I., Guelly, C., Trajanoski, S., and Puchhammer-Stöckl, E. (2010). Deep sequencing reveals highly complex dynamics of human cytomegalovirus genotypes in transplant patients over time. *Journal of Virology* 84, 7195–7203. doi:10.1128/JVI.00475-10.

Graci, J. D., and Cameron, C. E. (2005). Mechanisms of action of ribavirin against distinct viruses. *Rev. Med. Virol.* 16, 37–48. doi:10.1002/rmv.483.

Grande-Perez, A., Sierra, S., Castro, M. G., Domingo, E., and Lowenstein, P. R. (2002). Molecular indetermination in the transition to error catastrophe: systematic elimination of lymphocytic choriomeningitis virus through mutagenesis does not correlate linearly with large increases in mutant spectrum complexity. *Proc. Natl. Acad. Sci. U.S.A.* 99, 12938–12943. doi:10.1073/pnas.182426999.

Grande-Pérez, A., Gómez-Mariano, G., Lowenstein, P. R., and Domingo, E. (2005a). Mutagenesis-induced, large fitness variations with an invariant arenavirus consensus genomic nucleotide sequence. *Journal of Virology* 79, 10451–10459. doi:10.1128/JVI.79.16.10451-10459.2005.

Grande-Pérez, A., Lázaro, E., Lowenstein, P., Domingo, E., and Manrubia, S. C. (2005b). Suppression of viral infectivity through lethal defection. *Proc Natl Acad Sci USA* 102, 4448–4452. doi:10.1073/pnas.0408871102.

Gregori, J., Salicrú, M., Domingo, E., Sanchez, A., Esteban, J. I., Rodríguez-Frías, F., et al. (2014). Inference with viral quasispecies diversity indices: clonal and NGS approaches. *Bioinformatics* 30, 1104–1111. doi:10.1093/bioinformatics/btt768.

Gregori, J., Perales, C., Rodríguez-Frías, F., Esteban, J. I., Quer, J., and Domingo, E. (2016). Viral quasispecies complexity measures. *Virology* 493, 227–237. doi:10.1016/j.virol.2016.03.017.

Grigoras, I., Timchenko, T., Grande-Pérez, A., Katul, L., Vetten, H.-J., and Gronenborn, B. (2010). High variability and rapid evolution of a nanovirus. *Journal of Virology* 84, 9105–9117. doi:10.1128/JVI.00607-10.

Guenoune-Gelbart, D., Elbaum, M., Sagi, G., Levy, A., and Epel, B. L. (2008). Tobacco mosaic virus (TMV) replicase and movement protein function synergistically in facilitating TMV spread by lateral diffusion in the plasmodesmal desmotubule of *Nicotiana benthamiana*. *MPMI* 21, 335–345. doi:10.1094/MPMI-21-3-0335.

Guilliam, T. A., Jozwiakowski, S. K., Ehlinger, A., Barnes, R. P., Rudd, S. G., Bailey, L. J., et al. (2015). Human PrimPol is a highly error-prone polymerase regulated by single-stranded DNA binding proteins. *Nucleic Acids Research* 43, 1056–1068. doi:10.1093/nar/gku1321.

Guo, C., Fischhaber, P. L., Paszyc, M. J. L., Masuda, Y., Zhou, J., Kamiya, K., et al. (2003). Mouse Rev1 protein interacts with multiple DNA polymerases involved in translesion DNA synthesis. *EMBO J.* 22, 6621–6630. doi:10.1093/emboj/cdg626.

Guo, C., Sonoda, E., Tang, T.-S., Parker, J. L., Bielen, A. B., Takeda, S., et al. (2006). REV1 protein interacts with PCNA: significance of the REV1 BRCT domain in vitro and in vivo. *Mol. Cell* 23, 265–271. doi:10.1016/j.molcel.2006.05.038.

Gutierrez, C. (1999). Geminivirus DNA replication. *CMLS, Cell. Mol. Life Sci.* 56, 313–329. doi:10.1007/s000180050433.

Hammond-Kosack, K. and Jones, J.D., 2000. Responses to plant pathogens.

Biochemistry and molecular biology of plants, pp.1102-1156.

Hanley-Bowdoin, L., Bejarano, E. R., Robertson, D., and Mansoor, S. (2013). Geminiviruses: masters at redirecting and reprogramming plant processes. *Nat Rev Micro* 11, 777–788. doi:10.1038/nrmicro3117.

Hanley-Bowdoin, L., Settlage, S. B., and Robertson, D. (2004). Reprogramming plant gene expression: a prerequisite to geminivirus DNA replication. *Molecular Plant Pathology* 5, 149–156. doi:10.1111/j.1364-3703.2004.00214.x.

Haracska, L., Prakash, L., and Prakash, S. (2002). Role of human DNA polymerase kappa as an extender in translesion synthesis. *Proc Natl Acad Sci USA* 99, 16000–16005. doi:10.1073/pnas.252524999.

Haracska, L., Yu, S. L., Johnson, R. E., Prakash, L., and Prakash, S. (2000). Efficient and accurate replication in the presence of 7,8-dihydro-8-oxoguanine by DNA polymerase eta. *Nat. Genet.* 25, 458–461. doi:10.1038/78169.

Harrison, B. D., and Robinson, D. J. (1999). Natural genomic and antigenic variation in whitefly-transmitted geminiviruses (begomoviruses). *Annu Rev Phytopathol* 37, 369–398. doi:10.1146/annurev.phyto.37.1.369.

Hedskog, C., Mild, M., Jernberg, J., Sherwood, E., Bratt, G., Leitner, T., et al. (2010). Dynamics of HIV-1 quasispecies during antiviral treatment dissected using ultra-deep pyrosequencing. *PLoS ONE* 5, e11345. doi:10.1371/journal.pone.0011345.

Heinlein, M. (2015). Plant virus replication and movement. *Virology* 479-480, 657–671. doi:10.1016/j.virol.2015.01.025.

Hendel, A., Krijger, P. H. L., Diamant, N., Goren, Z., Langerak, P., Kim, J., et al. (2011). PCNA Ubiquitination Is Important, But Not Essential for Translesion DNA Synthesis in Mammalian Cells. *PLoS Genet* 7, e1002262–13. doi:10.1371/journal.pgen.1002262.

Henn, M. R., Boutwell, C. L., Charlebois, P., Lennon, N. J., Power, K. A., Macalalad, A. R., et al. (2012). Whole genome deep sequencing of HIV-1 reveals the impact of early minor variants upon immune recognition during acute infection. *PLoS Pathog* 8, e1002529. doi:10.1371/journal.ppat.1002529.

Hicks, A. L., and Duffy, S. (2014). Cell tropism predicts long-term nucleotide substitution rates of mammalian RNA viruses. *PLoS Pathog* 10, e1003838. doi:10.1371/journal.ppat.1003838.

Hicks, C., Clay, P., Redfield, R., Lalezari, J., Liporace, R., Schneider, S., et al. (2013). Safety, tolerability, and efficacy of KP-1461 as monotherapy for 124 days in antiretroviral-experienced, HIV type 1-infected subjects. *AIDS Res. Hum. Retroviruses* 29, 250–255. doi:10.1089/AID.2012.0093.

Hino, S., and Miyata, H. (2007). Torque teno virus (TTV): current status. *Rev. Med. Virol.* 17, 45–57. doi:10.1002/rmv.524.

- Hogenhout, S. A., Ammar, E.-D., Whitfield, A. E., and Redinbaugh, M. G. (2008). Insect vector interactions with persistently transmitted viruses. *Annu Rev Phytopathol* 46, 327–359. doi:10.1146/annurev.phyto.022508.092135.
- Holland, J. J., Domingo, E., la Torre, de, J. C., and Steinhauer, D. A. (1990). Mutation frequencies at defined single codon sites in vesicular stomatitis virus and poliovirus can be increased only slightly by chemical mutagenesis. *Journal of Virology* 64, 3960–3962.
- Holland, J., Spindler, K., Horodyski, F., Grabau, E., Nichol, S., and VandePol, S. (1982). Rapid evolution of RNA genomes. *Science* 215, 1577–1585.
- Holley, R. W., Apgar, J., Everett, G. A., Madison, J. T., Marquisee, M., Merrill, S. H., et al. (1965). Structure of a ribonucleic acid. *Science* 147, 1462–1465.
- Holmes, F. O. (1955). Preventive and curative effects of thiouracil treatments in mosaichypersensitive tobacco. *Virology* 1, 1–9.
- Hormuzdi, S. G., and Bisaro, D. M. (1995). Genetic analysis of beet curly top virus: examination of the roles of L2 and L3 genes in viral pathogenesis. *Virology* 206, 1044–1054. doi:10.1006/viro.1995.1027.
- Horner, D. S., Pavesi, G., Castrignanò, T., De Meo, P. D., Liuni, S., Sammeth, M., et al. (2010). Bioinformatics approaches for genomics and post genomics applications of next-generation sequencing. *Brief Bioinform* 11, 181–197. doi:10.1093/bib/bbp046.
- Hou, W., and Bonkovsky, H. L. (2013). Non-coding RNAs in hepatitis C-induced hepatocellular carcinoma: dysregulation and implications for early detection, diagnosis and therapy. *World J. Gastroenterol.* 19, 7836–7845. doi:10.3748/wjg.v19.i44.7836.
- Hou, Y. M., and Gilbertson, R. L. (1996). Increased pathogenicity in a pseudorecombinant bipartite geminivirus correlates with intermolecular recombination. *Journal of Virology* 70, 5430–5436.
- Howorka, S., Cheley, S., and Bayley, H. (2001). Sequence-specific detection of individual DNA strands using engineered nanopores. *Nat. Biotechnol.* 19, 636–639. doi:10.1038/90236.
- Höhnle, M., Höfer, P., Bedford, I. D., Briddon, R. W., Markham, P. G., and Frischmuth, T. (2001). Exchange of three amino acids in the coat protein results in efficient whitefly transmission of a nontransmissible Abutilon mosaic virus isolate. *Virology* 290, 164–171. doi:10.1006/viro.2001.1140.
- Hu, J. S., Gelman, D. B., and Blackburn, M. B. (2003). Age-specific interaction between the parasitoid, *Encarsia formosa* and its host, the silverleaf whitefly, *Bemisia tabaci* (Strain B). *J. Insect Sci.* 3, 28.
- Hull, R. (2013). *Plant Virology*. Academic Press.
- Idris, A. M., and Brown, J. K. (2005). Evidence for interspecific-recombination for three monopartite begomoviral genomes associated with the tomato leaf curl disease

from central Sudan. *Arch Virol* 150, 1003–1012. doi:10.1007/s00705-004-0484-7.

Iranzo, J., and Manrubia, S. C. (2009). Stochastic extinction of viral infectivity through the action of defectors. *EPL* 85, 18001. doi:10.1209/0295-5075/85/18001.

Isnard, M., Granier, M., Frutos, R., Reynaud, B., and Peterschmitt, M. (1998). Quasispecies nature of three maize streak virus isolates obtained through different modes of selection from a population used to assess response to infection of maize cultivars. *Journal of General Virology* 79 (Pt 12), 3091–3099. doi:10.1099/0022-1317-79-12-3091.

Ivanowski, D. (1968). Concerning the mosaic disease of the tobacco plant.

Jansen, J. G., Langerak, P., Tsaalbi-Shtylik, A., van den Berk, P., Jacobs, H., and de Wind, N. (2006). Strand-biased defect in C/G transversions in hypermutating immunoglobulin genes in Rev1-deficient mice. *J. Exp. Med.* 203, 319–323. doi:10.1084/jem.20052227.

Jesús Santiago, M., Alejandro-Durán, E., Muñoz-Serrano, A., and Ruiz-Rubio, M. (2008). Two translesion synthesis DNA polymerase genes, AtPOLH and AtREV1, are involved in development and UV light resistance in Arabidopsis. *J. Plant Physiol.* 165, 1582–1591. doi:10.1016/j.jplph.2007.11.012.

Johnson, R. E., Kondratyck, C. M., Prakash, S., and Prakash, L. (1999). hRAD30 mutations in the variant form of xeroderma pigmentosum. *Science* 285, 263–265.

Johnson, R. E., Prakash, S., and Prakash, L. (2000). The human DINB1 gene encodes the DNA polymerase Poltheta. *Proc Natl Acad Sci USA* 97, 3838–3843.

Jupin, I., De Kouchkovsky, F., Jouanneau, F., and Gronenborn, B. (1994). Movement of tomato yellow leaf curl geminivirus (TYLCV): involvement of the protein encoded by ORF C4. *Virology* 204, 82–90. doi:10.1006/viro.1994.1512.

Kaliappan, K., Choudhury, N. R., Suyal, G., and Mukherjee, S. K. (2012). A novel role for RAD54: this host protein modulates geminiviral DNA replication. *The FASEB Journal* 26, 1142–1160. doi:10.1096/fj.11-188508.

Kanao, R., Masuda, Y., Deguchi, S., Yumoto-Sugimoto, M., Hanaoka, F., and Masutani, C. (2015). Relevance of simultaneous mono-ubiquitinations of multiple units of PCNA homo-trimers in DNA damage tolerance. *PLoS ONE* 10, e0118775. doi:10.1371/journal.pone.0118775.

Kang, B.-C., Yeam, I., and Jahn, M. M. (2005). Genetics of plant virus resistance. *Annu Rev Phytopathol* 43, 581–621. doi:10.1146/annurev.phyto.43.011205.141140.

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16, 111–120.

Kimura, S., and Sakaguchi, K. (2006). DNA repair in plants. *Chem. Rev.* 106, 753–766.

doi:10.1021/cr040482n.

Kirthi, N., Maiya, S. P., Murthy, M. R. N., and Savithri, H. S. (2002). Evidence for recombination among the tomato leaf curl virus strains/species from Bangalore, India. *Arch Virol* 147, 255–272.

Knipe, D. M., Howley, P. M., Griffin, D., Lamb, R., and Martin, M. (2001). *Fields virology*, vol. 1. ... (EUA): Lippincott Williams &

Knoepfel, S. A., Di Giallonardo, F., Däumer, M., Thielen, A., and Metzner, K. J. (2011). In-depth analysis of G-to-A hypermutation rate in HIV-1 env DNA induced by endogenous APOBEC3 proteins using massively parallel sequencing. *Journal of Virological Methods* 171, 329–338. doi:10.1016/j.jviromet.2010.11.016.

Kong, L.-J., and Hanley-Bowdoin, L. (2002). A geminivirus replication protein interacts with a protein kinase and a motor protein that display different expression patterns during plant development and infection. *The Plant Cell* 14, 1817–1832. doi:10.1105/tpc.003681.

Kunz, B. A. (2008). A role for PCNA2 in translesion synthesis by *Arabidopsis thaliana* DNA polymerase ϵ . *Plant Signal Behav* 3, 897–898. doi:10.1111/j.1365-313X.2008.03562.x.

Kunz, B. A., Anderson, H. J., Osmond, M. J., and Vonarx, E. J. (2005). Components of nucleotide excision repair and DNA damage tolerance in *Arabidopsis thaliana*. *Environ. Mol. Mutagen.* 45, 115–127. doi:10.1002/em.20094.

Kusov, Y., Tan, J., Alvarez, E., Enjuanes, L., and Hilgenfeld, R. (2015). A G-quadruplex-binding macrodomain within the “SARS-unique domain” is essential for the activity of the SARS-coronavirus replication-transcription complex. *Virology* 484, 313–322. doi:10.1016/j.virol.2015.06.016.

Kutnjak, D., Rupar, M., Gutierrez-Aguirre, I., Curk, T., Kreuze, J. F., and Ravnkar, M. (2015). Deep Sequencing of Virus-Derived Small Interfering RNAs and RNA from Viral Particles Shows Highly Similar Mutational Landscapes of a Plant Virus Population. *Journal of Virology* 89, 4760–4769. doi:10.1128/JVI.03685-14.

la Torre, de, J. C., and Holland, J. J. (1990). RNA virus quasispecies populations can suppress vastly superior mutant progeny. *Journal of Virology* 64, 6278–6281.

Ladoukakis, E. D., Theologidis, I., Rodakis, G. C., and Zouros, E. (2011). Homologous recombination between highly diverged mitochondrial sequences: examples from maternally and paternally transmitted genomes. *Mol. Biol. Evol.* 28, 1847–1859. doi:10.1093/molbev/msr007.

Lam, E. Y. N., Beraldi, D., Tannahill, D., and Balasubramanian, S. (2013). G-quadruplex structures are stable and detectable in human genomic DNA. *Nat Commun* 4, 1796–16. doi:10.1038/ncomms2792.

Lataillade, M., Chiarella, J., Yang, R., DeGrosky, M., Uy, J., Seekins, D., et al. (2012). Virologic failures on initial boosted-PI regimen infrequently possess low-level variants with major PI resistance mutations by ultra-deep sequencing. *PLoS ONE* 7, e30118. doi:10.1371/journal.pone.0030118.

Laufs, J., Jupin, I., David, C., Schumacher, S., Heyraud-Nitschke, F., and Gronenborn, B. (1995a). Geminivirus replication: genetic and biochemical characterization of Rep protein function, a review. *Biochimie* 77, 765–773.

Laufs, J., Traut, W., Heyraud, F., Matzeit, V., Rogers, S. G., Schell, J., et al. (1995b). In vitro cleavage and joining at the viral origin of replication by the replication initiator protein of tomato yellow leaf curl virus. *Proc Natl Acad Sci USA* 92, 3879–3883.

Lee, G. Q., Swenson, L. C., Poon, A. F. Y., Martin, J. N., Hatano, H., Deeks, S. G., et al. (2012). Prolonged and substantial discordance in prevalence of raltegravir-resistant HIV-1 in plasma versus PBMC samples revealed by 454 “deep” sequencing. *PLoS ONE* 7, e46181. doi:10.1371/journal.pone.0046181.

Lefevre, P., Lett, J.-M., Reynaud, B., and Martin, D. P. (2007). Avoidance of protein fold disruption in natural virus recombinants. *PLoS Pathog* 3, e181. doi:10.1371/journal.ppat.0030181.

Lefevre, P., Martin, D. P., Harkins, G., Lemey, P., Gray, A. J. A., Meredith, S., et al. (2010). The Spread of Tomato Yellow Leaf Curl Virus from the Middle East to the World. *PLoS Pathog* 6, e1001164–12. doi:10.1371/journal.ppat.1001164.

Lehmann, A. R. (2002). Replication of damaged DNA in mammalian cells: new solutions to an old problem. *Mutat. Res.* 509, 23–34.

Lehmann, A. R., McGibbon, D., and Stefanini, M. (2011). Xeroderma pigmentosum. *Orphanet J Rare Dis* 6, 70. doi:10.1186/1750-1172-6-70.

Lehmann, A. R., Niimi, A., Ogi, T., Brown, S., Sabbioneda, S., Wing, J. F., et al. (2007). Translesion synthesis: Y-family polymerases and the polymerase switch. *DNA Repair (Amst.)* 6, 891–899. doi:10.1016/j.dnarep.2007.02.003.

Lemontt, J. F. (1971). Pathways of ultraviolet mutability in *Saccharomyces cerevisiae*. I. Some properties of double mutants involving *uvr9* and *rev*. *Mutat. Res.* 13, 311–317.

Lewandowski, D. J., and Dawson, W. O. (1999). *Tobamoviruses*. Encyclopedia of virology.

Li, W.-H. (1993). Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *J. Mol. Evol.* 36, 96–99. doi:10.1007/BF02407308.

Ling, H., Boudsocq, F., Woodgate, R., and Yang, W. (2001). Crystal structure of a Y-family DNA polymerase in action: a mechanism for error-prone and lesion-bypass replication. *Cell* 107, 91–102.

Lisha, V. S., Antony, B., Palaniswami, M. S., and Henneberry, T. J. (2003). *Bemisia tabaci* (Homoptera: Aleyrodidae) biotypes in India. *J. Econ. Entomol.* 96, 322–327.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., et al. (2012). Comparison of Next-Generation Sequencing Systems. *Journal of Biomedicine and Biotechnology* 2012, 1–11. doi:10.1155/2012/251364.

Loeb, L. A., Essigmann, J. M., Kazazi, F., Zhang, J., Rose, K. D., and Mullins, J. I. (1999). Lethal mutagenesis of HIV with mutagenic nucleoside analogs. *Proc Natl Acad Sci USA* 96, 1492–1497.

Loman, N. J., Misra, R. V., Dallman, T. J., Constantinidou, C., Gharbia, S. E., Wain, J., et al. (2012). Performance comparison of benchtop high-throughput sequencing platforms. *Nat. Biotechnol.* 30, 434–439. doi:10.1038/nbt.2198.

Longley, D. B., Harkin, D. P., and Johnston, P. G. (2003). 5-Fluorouracil: mechanisms of action and clinical strategies. *Nat Rev Cancer* 3, 330–338. doi:10.1038/nrc1074.

López-Bueno, A., Mateu, M. G., and Almendral, J. M. (2003). High mutant frequency in populations of a DNA virus allows evasion from antibody therapy in an immunodeficient host. *Journal of Virology* 77, 2701–2708. doi:10.1128/JVI.77.4.2701-2708.2003.

López-Bueno, A., Villarreal, L. P., and Almendral, J. M. (2006). Parvovirus variation for disease: a difference with RNA viruses? *Curr. Top. Microbiol. Immunol.* 299, 349–370.

Lucas, W. J. (2006). Plant viral movement proteins: agents for cell-to-cell trafficking of viral genomes. *Virology* 344, 169–184. doi:10.1016/j.virol.2005.09.026.

Luque, A., Sanz-Burgos, A. P., Ramirez-Parra, E., Castellano, M. M., and Gutierrez, C. (2002). Interaction of geminivirus Rep protein with replication factor C and its potential role during geminivirus DNA replication. *Virology* 302, 83–94.

Lutchman, G., Danehower, S., Song, B.-C., Liang, T. J., Hoofnagle, J. H., Thomson, M., et al. (2007). Mutation rate of the hepatitis C virus NS5B in patients undergoing treatment with ribavirin monotherapy. *Gastroenterology* 132, 1757–1766. doi:10.1053/j.gastro.2007.03.035.

Maag, D., Castro, C., Hong, Z., and Cameron, C. E. (2001). Hepatitis C virus RNA-dependent RNA polymerase (NS5B) as a mediator of the antiviral activity of ribavirin. *Journal of Biological Chemistry* 276, 46094–46098. doi:10.1074/jbc.C100349200.

MacFarlane, S. A. (1997). Natural Recombination among Plant Virus Genomes: Evidence from Tobraviruses. *Seminars in Virology* 8, 25–31. doi:10.1006/smvv.1997.0103.

Maizels, N. (2015). G4-associated human diseases. *EMBO Rep.* doi:10.15252/embr.201540607.

Malik, P. S., Kumar, V., Bagewadi, B., and Mukherjee, S. K. (2005). Interaction

between coat protein and replication initiation protein of Mung bean yellow mosaic India virus might lead to control of viral DNA replication. *Virology* 337, 273–283. doi:10.1016/j.virol.2005.04.030.

Malpica, J. M., Fraile, A., Moreno, I., Obies, C. I., Drake, J. W., and García-Arenal, F. (2002). The rate and character of spontaneous mutation in an RNA virus. *Genetics* 162, 1505–1511.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376–380. doi:10.1038/nature03959.

Martin, D., and Rybicki, E. (2000). RDP: detection of recombination amongst aligned sequences. *Bioinformatics* 16, 562–563.

Martín, V., Abia, D., Domingo, E., and Grande-Pérez, A. (2010). An interfering activity against lymphocytic choriomeningitis virus replication associated with enhanced mutagenesis. *J. Gen. Virol.* 91, 990–1003. doi:10.1099/vir.0.017053-0.

Martín, V., Perales, C., Dávila, M., and Domingo, E. (2006). Viral fitness can influence the repertoire of virus variants selected by antibodies. *Journal of Molecular Biology* 362, 44–54. doi:10.1016/j.jmb.2006.06.077.

Martínez, F., Lafforgue, G., Morelli, M. J., González-Candelas, F., Chua, N.-H., Daròs, J.-A., et al. (2012). Ultradeep sequencing analysis of population dynamics of virus escape mutants in RNAi-mediated resistant plants. *Mol. Biol. Evol.* 29, 3297–3307. doi:10.1093/molbev/mss135.

Martínez-Jiménez, M. I., García-Gómez, S., Bebenek, K., Sastre-Moreno, G., Calvo, P. A., Díaz-Talavera, A., et al. (2015). Alternative solutions and new scenarios for translesion DNA synthesis by human PrimPol. *DNA Repair (Amst.)* 29, 127–138. doi:10.1016/j.dnarep.2015.02.013.

Más, P., and Beachy, R. N. (1999). Replication of tobacco mosaic virus on endoplasmic reticulum and role of the cytoskeleton and virus movement protein in intracellular distribution of viral RNA. *J Cell Biol* 147, 945–958.

Mason, G., Caciagli, P., Accotto, G. P., and Noris, E. (2008). Real-time PCR for the quantitation of Tomato yellow leaf curl Sardinia virus in tomato plants and in *Bemisia tabaci*. *Journal of Virological Methods* 147, 282–289. doi:10.1016/j.jviromet.2007.09.015.

Massart, S., Olmos, A., Jijakli, H., and Candresse, T. (2014). Current impact and future directions of high throughput sequencing in plant virus diagnostics. *Virus Research* 188, 90–96. doi:10.1016/j.virusres.2014.03.029.

Masuda, K., Ouchida, R., Li, Y., Gao, X., Mori, H., and Wang, J.-Y. (2009). A critical role for REV1 in regulating the induction of C:G transitions and A:T mutations during Ig gene hypermutation. *J. Immunol.* 183, 1846–1850. doi:10.4049/jimmunol.0901240.

Masuda, Y., Kanao, R., Kaji, K., Ohmori, H., Hanaoka, F., and Masutani, C. (2015). Different types of interaction between PCNA and PIP boxes contribute to distinct cellular functions of Y-family DNA polymerases. *Nucleic Acids Research* 43, 7898–7910. doi:10.1093/nar/gkv712.

Masutani, C., Kusumoto, R., Yamada, A., Dohmae, N., Yokoi, M., Yuasa, M., et al. (1999). The XPV (xeroderma pigmentosum variant) gene encodes human DNA polymerase η . *Nature* 399, 700–704. doi:10.1038/21447.

Matsuda, T., Bebenek, K., Masutani, C., Hanaoka, F., and Kunkel, T. A. (2000). Low fidelity DNA synthesis by human DNA polymerase- η . *Nature* 404, 1011–1013. doi:10.1038/35010014.

Mayer, A. (1886). *Ueber die Mosaikkrankheit des Tabaks*. Die Landwirtschaftlichen Versuchs-Stationen.

Más, P., and Beachy, R. N. (1999). Replication of tobacco mosaic virus on endoplasmic reticulum and role of the cytoskeleton and virus movement protein in intracellular distribution of viral RNA. *J Cell Biol* 147, 945–958. doi:10.1083/jcb.147.5.945.

McCulloch, S. D., and Kunkel, T. A. (2008). The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Nature Publishing Group* 18, 148–161. doi:10.1038/cr.2008.4.

McIlwraith, M. J., McIlwraith, M. J., Vaisman, A., Liu, Y., Fanning, E., Woodgate, R., et al. (2005). Human DNA polymerase η promotes DNA synthesis from strand invasion intermediates of homologous recombination. *Mol. Cell* 20, 783–792. doi:10.1016/j.molcel.2005.10.001.

Meshi, T., Watanabe, Y., Saito, T., Sugimoto, A., Maeda, T., and Okada, Y. (1987). Function of the 30 kd protein of tobacco mosaic virus: involvement in cell-to-cell movement and dispensability for replication. *EMBO J.* 6, 2557–2563.

Métifiot, M., Amrane, S., Litvak, S., and Andreola, M.-L. (2014). G-quadruplexes in viruses: function and potential therapeutic applications. *Nucleic Acids Research* 42, 12352–12366. doi:10.1093/nar/gku999.

Mild, M., Hedskog, C., Jernberg, J., and Albert, J. (2011). Performance of ultra-deep pyrosequencing in analysis of HIV-1 pol gene variation. *PLoS ONE* 6, e22741. doi:10.1371/journal.pone.0022741.

Miller, J. R., Koren, S., and Sutton, G. (2010). Assembly algorithms for next-generation sequencing data. *Genomics* 95, 315–327. doi:10.1016/j.ygeno.2010.03.001.

Minskaia, E., Hertzog, T., Gorbalenya, A. E., Campanacci, V., Cambillau, C., Canard, B., et al. (2006). Discovery of an RNA virus 3'→5' exoribonuclease that is critically involved in coronavirus RNA synthesis. *Proc Natl Acad Sci USA* 103, 5108–5113. doi:10.1073/pnas.0508200103.

Mitsuya, Y., Varghese, V., Wang, C., Liu, T. F., Holmes, S. P., Jayakumar, P., et al. (2008). Minority human immunodeficiency virus type 1 variants in antiretroviral-naïve persons with reverse transcriptase codon 215 revertant mutations. *Journal of Virology* 82, 10747–10755. doi:10.1128/JVI.01827-07.

Monci, F., Sánchez-Campos, S., Navas-Castillo, J., and Moriones, E. (2002a). A natural recombinant between the geminiviruses Tomato yellow leaf curl Sardinia virus and Tomato yellow leaf curl virus exhibits a novel pathogenic phenotype and is becoming prevalent in Spanish populations. *Virology* 303, 317–326.

Monci, F., Sánchez-Campos, S., Navas-Castillo, J., and Moriones, E. (2002b). A natural recombinant between the geminiviruses Tomato yellow leaf curl Sardinia virus and Tomato yellow leaf curl virus exhibits a novel pathogenic phenotype and is becoming prevalent in Spanish populations. *Virology* 303, 317–326. doi:10.1006/viro.2002.1633.

Montarry, J., DOUMAYROU, J., Simon, V., and Moury, B. (2011). Genetic background matters: a plant–virus gene-for-gene interaction is strongly influenced by genetic contexts. *Molecular Plant Pathology* 12, 911–920. doi:10.1111/j.1364-3703.2011.00724.x.

Moreira, D., and López-García, P. (2009). Ten reasons to exclude viruses from the tree of life. *Nat Rev Micro* 7, 306–311. doi:10.1038/nrmicro2108.

Moreno, E., Ojosnegros, S., García-Arriaza, J., Escarmís, C., Domingo, E., and Perales, C. (2014). Exploration of sequence space as the basis of viral RNA genome segmentation. *Proc. Natl. Acad. Sci. U.S.A.* 111, 6678–6683. doi:10.1073/pnas.1323136111.

Moreno, H., Gallego, I., Sevilla, N., la Torre, de, J. C., Domingo, E., and Martín, V. (2011). Ribavirin can be mutagenic for arenaviruses. *Journal of Virology* 85, 7246–7255. doi:10.1128/JVI.00614-11.

Moreno, H., Grande-Pérez, A., Domingo, E., and Martín, V. (2012). Arenaviruses and Lethal Mutagenesis. Prospects for New Ribavirin-based Interventions. *Viruses* 4, 2786–2805. doi:10.3390/v4112786.

Moreno, I. M., Malpica, J. M., Díaz-Pendón, J. A., Moriones, E., Fraile, A., and García-Arenal, F. (2004). Variability and genetic structure of the population of watermelon mosaic virus infecting melon in Spain. *Virology* 318, 451–460. doi:10.1016/j.virol.2003.10.002.

Moreno, I. M., Malpica, J. M., Rodríguez-Cerezo, E., and García-Arenal, F. (1997). A mutation in tomato aspermy cucumovirus that abolishes cell-to-cell movement is maintained to high levels in the viral RNA population by complementation. *Journal of Virology* 71, 9157–9162.

Morilla, G., Antúnez, C., Bejarano, E. R., Janssen, D., and Cuadrado, I. M. (2007). A New Tomato yellow leaf curl virus Strain in Southern Spain. <http://dx.doi.org/10.1094/PDIS.2003.87.8.1004B> 87, 1004–1004.

doi:10.1094/PDIS.2003.87.8.1004B.

Morilla, G., Castillo, A. G., Preiss, W., Jeske, H., and Bejarano, E. R. (2006). A Versatile Transreplication-Based System To Identify Cellular Proteins Involved in Geminivirus Replication. *Journal of Virology* 80, 3624–3633. doi:10.1128/JVI.80.7.3624-3633.2006.

Morilla, G., Krenz, B., Jeske, H., Bejarano, E. R., and Wege, C. (2004). Tête à tête of tomato yellow leaf curl virus and tomato yellow leaf curl sardinia virus in single nuclei. *Journal of Virology* 78, 10715–10723. doi:10.1128/JVI.78.19.10715-10723.2004.

Moriones, E., and Navas-Castillo, J. (2000). Tomato yellow leaf curl virus, an emerging virus complex causing epidemics worldwide. *Virus Research* 71, 123–134. doi:10.1016/S0168-1702(00)00193-3.

Morroni, M., Jacquemond, M., and Tepfer, M. (2013). Deep Sequencing of Recombinant Virus Populations in Transgenic and Nontransgenic Plants Infected with Cucumber mosaic virus. <http://dx.doi.org/10.1094/MPMI-02-13-0057-R> 26, 801–811. doi:10.1094/MPMI-02-13-0057-R.

Mortier, V., Dauwe, K., Vancoillie, L., Staelens, D., Van Wanzeele, F., Vogelaers, D., et al. (2013). Frequency and predictors of HIV-1 co-receptor switch in treatment naive patients. *PLoS ONE* 8, e80259. doi:10.1371/journal.pone.0080259.

Mullins, J. I., Heath, L., Hughes, J. P., Kicha, J., Styrchak, S., Wong, K. G., et al. (2011). Mutation of HIV-1 Genomes in a Clinical Population Treated with the Mutagenic Nucleoside KP1461. *PLoS ONE* 6, e15135. doi:10.1371/journal.pone.0015135.

Mundry, K. W., Watkins, P. A., Ashfield, T., Plaskitt, K. A., Eisele-Walter, S., and Wilson, T. M. (1991). Complete uncoating of the 5' leader sequence of tobacco mosaic virus RNA occurs rapidly and is required to initiate cotranslational virus disassembly in vitro. *Journal of General Virology* 72 (Pt 4), 769–777. doi:10.1099/0022-1317-72-4-769.

Murakumo, Y., Ogura, Y., Ishii, H., Numata, S., Ichihara, M., Croce, C. M., et al. (2001). Interactions in the error-prone postreplication repair proteins hREV1, hREV3, and hREV7. *Journal of Biological Chemistry* 276, 35644–35651. doi:10.1074/jbc.M102051200.

Murat, P., Zhong, J., Lekieffre, L., Cowieson, N. P., Clancy, J. L., Preiss, T., et al. (2014). G-quadruplexes regulate Epstein-Barr virus-encoded nuclear antigen 1 mRNA translation. *Nat Chem Biol* 10, 358–364. doi:10.1038/nchembio.1479.

Nagy, P. D., and Simon, A. E. (1997). New insights into the mechanisms of RNA recombination. *Virology* 235, 1–9. doi:10.1006/viro.1997.8681.

Nair, D. T., Johnson, R. E., Prakash, L., and Prakash, S. (2005). Rev1 employs a novel mechanism of DNA synthesis using a protein template. *Science*. doi:10.1126/science.1116510.

Nakagawa, M., Takahashi, S., Narumi, I., and Sakamoto, A. N. (2011a). Role of AtPolζ,

AtRev1 and AtPol η in γ ray-induced mutagenesis. *Plant Signal Behav* 6, 728–731. doi:10.4161/psb.6.5.15124.

Nakagawa, M., Takahashi, S., Tanaka, A., Narumi, I., and Sakamoto, A. N. (2011b). Role of AtPol ζ , AtRev1, and AtPol η in UV light-induced mutagenesis in Arabidopsis. *PLANT PHYSIOLOGY* 155, 414–420. doi:10.1104/pp.110.166082.

Navas-Castillo, J., Sánchez-Campos, S., and Díaz, J. A. (1999). Tomato yellow leaf curl virus-Is causes a novel disease of common bean and severe epidemics in tomato in Spain. *Plant* doi:10.1094/PDIS.1999.83.1.29.

Navas-Castillo, J., Sánchez-Campos, S., Noris, E., Louro, D., Accotto, G. P., and Moriones, E. (2000). Natural recombination between Tomato yellow leaf curl virus-is and Tomato leaf curl virus. *Journal of General Virology* 81, 2797–2801. doi:10.1099/0022-1317-81-11-2797.

Ndunguru, J., Legg, J. P., Aveling, T. A. S., Thompson, G., and Fauquet, C. M. (2005). Molecular biodiversity of cassava begomoviruses in Tanzania: evolution of cassava geminiviruses in Africa and evidence for East Africa being a center of diversity of cassava geminiviruses. *Virology Journal* 2, 21. doi:10.1186/1743-422X-2-21.

Nelson, J. R., Lawrence, C. W., and Hinkle, D. C. (1996). Deoxycytidyl transferase activity of yeast REV1 protein. *Nature* 382, 729–731. doi:10.1038/382729a0.

Niehl, A., Peña, E. J., Amari, K., and Heinlein, M. (2013). Microtubules in viral replication and transport. *The Plant Journal* 75, 290–308. doi:10.1111/tpj.12134.

Ninomiya, M., Ueno, Y., Funayama, R., Nagashima, T., Nishida, Y., Kondo, Y., et al. (2012). Use of illumina deep sequencing technology to differentiate hepatitis C virus variants. *Journal of Clinical Microbiology* 50, 857–866. doi:10.1128/JCM.05715-11.

Nishiguchi, M., and Motoyoshi, F. (1978). Behaviour of a temperature sensitive strain of tobacco mosaic virus in tomato leaves and protoplasts. *Journal of General*

Noguchi, H., Park, J., and Takagi, T. (2006). MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. *Nucleic Acids Research* 34, 5623–5630. doi:10.1093/nar/gkl723.

Noris, E., Vaira, A. M., Caciagli, P., Masenga, V., Gronenborn, B., and Accotto, G. P. (1998). Amino acids in the capsid protein of tomato yellow leaf curl virus that are crucial for systemic infection, particle formation, and insect transmission. *Journal of Virology* 72, 10050–10057.

Ogi, T., Kato, T., and Ohmori, H. (1999). Mutation enhancement by DINB1, a mammalian homologue of the Escherichia coli mutagenesis protein dinB. *Genes Cells* 4, 607–618.

Ohashi, E., Bebenek, K., Matsuda, T., Feaver, W. J., Gerlach, V. L., Friedberg, E. C., et al. (2000). Fidelity and processivity of DNA synthesis by DNA polymerase kappa, the

product of the human DINB1 gene. *Journal of Biological Chemistry* 275, 39678–39684. doi:10.1074/jbc.M005309200.

Ohashi, E., Murakumo, Y., Kanjo, N., Akagi, J.-I., Masutani, C., Hanaoka, F., et al. (2004). Interaction of hREV1 with three human Y-family DNA polymerases. *Genes Cells* 9, 523–531. doi:10.1111/j.1356-9597.2004.00747.x.

Ohmori, H., Friedberg, E. C., Fuchs, R. P., Goodman, M. F., Hanaoka, F., Hinkle, D., et al. (2001). The Y-family of DNA polymerases. *Mol. Cell* 8, 7–8.

Ojosnegros, S., García-Arriaza, J., Escarmís, C., Manrubia, S. C., Perales, C., Arias, A., et al. (2011). Viral genome segmentation can result from a trade-off between genetic content and particle stability. *PLoS Genet* 7, e1001344. doi:10.1371/journal.pgen.1001344.

Orozco, B. M., and Hanley-Bowdoin, L. (1996). A DNA structure is required for geminivirus replication origin function. *Journal of Virology* 70, 148–158.

Ortega-Prieto, A. M., Sheldon, J., Grande-Pérez, A., Tejero, H., Gregori, J., Quer, J., et al. (2013). Extinction of hepatitis C virus by ribavirin in hepatoma cells involves lethal mutagenesis. *PLoS ONE* 8, e71039. doi:10.1371/journal.pone.0071039.

Otsuka, C., Kunitomi, N., Iwai, S., Loakes, D., and Negishi, K. (2005). Roles of the polymerase and BRCT domains of Rev1 protein in translesion DNA synthesis in yeast in vivo. *Mutat. Res.* 578, 79–87. doi:10.1016/j.mrfmmm.2005.03.005.

Pabinger, S., Dander, A., Fischer, M., Snajder, R., Sperk, M., Efremova, M., et al. (2014). A survey of tools for variant analysis of next-generation genome sequencing data. *Brief Bioinform* 15, 256–278. doi:10.1093/bib/bbs086.

Padidam, M., Sawyer, S., and Fauquet, C. M. (1999). Possible emergence of new geminiviruses by frequent recombination. *Virology* 265, 218–225. doi:10.1006/viro.1999.0056.

Pagán, I., del Carmen Córdoba-Sellés, M., Martínez-Priego, L., Fraile, A., Malpica, J. M., Jordá, C., et al. (2006). Genetic Structure of the Population of Pepino mosaic virus Infecting Tomato Crops in Spain. *Phytopathology* 96, 274–279. doi:10.1094/PHYTO-96-0274.

Pamilo, P., and Bianchi, N. O. (1993). Evolution of the Zfx and Zfy genes: rates and interdependence between the genes. *Mol. Biol. Evol.* 10, 271–281.

Pariente, N., Airaksinen, A., and Domingo, E. (2003). Mutagenesis versus inhibition in the efficiency of extinction of foot-and-mouth disease virus. *Journal of Virology* 77, 7131–7138. doi:10.1128/JVI.77.12.7131-7138.2003.

Paszkiwicz, K., and Studholme, D. J. (2010). De novo assembly of short sequence reads. *Brief Bioinform* 11, bbq020–472. doi:10.1093/bib/bbq020.

Pathak, V. K., and Temin, H. M. (1992). 5-Azacytidine and RNA secondary structure

increase the retrovirus mutation rate. *Journal of Virology* 66, 3093–3100.

Patil, B. L., Dutt, N., Briddon, R. W., Bull, S. E., Rothenstein, D., Borah, B. K., et al. (2007). Deletion and recombination events between the DNA-A and DNA-B components of Indian cassava-infecting geminiviruses generate defective molecules in *Nicotiana benthamiana*. *Virus Research* 124, 59–67. doi:10.1016/j.virusres.2006.10.003.

Pelham, H. R. (1978). Leaky UAG termination codon in tobacco mosaic virus RNA. *Nature* 272, 469–471.

Perales, C., Agudo, R., Tejero, H., Manrubia, S. C., and Domingo, E. (2009). Potential Benefits of Sequential Inhibitor-Mutagen Treatments of RNA Virus Infections. *PLoS Pathog* 5, e1000658. doi:10.1371/journal.ppat.1000658.

Perales, C., Iranzo, J., Manrubia, S. C., and Domingo, E. (2012). The impact of quasispecies dynamics on the use of therapeutics. *Trends in Microbiology* 20, 595–603. doi:10.1016/j.tim.2012.08.010.

Perales, C., Martín, V., Ruiz-Jarabo, C. M., and Domingo, E. (2005). Monitoring sequence space as a test for the target of selection in viruses. *Journal of Molecular Biology* 345, 451–459. doi:10.1016/j.jmb.2004.10.066.

Perales, C., Mateo, R., Mateu, M. G., and Domingo, E. (2007). Insights into RNA virus mutant spectrum and lethal mutagenesis events: replicative interference and complementation by multiple point mutants. *Journal of Molecular Biology* 369, 985–1000. doi:10.1016/j.jmb.2007.03.074.

Perelson, A. S., and Layden, T. J. (2007). Ribavirin: is it a mutagen for hepatitis C virus? *Gastroenterology* 132, 2050–2052. doi:10.1053/j.gastro.2007.03.077.

Perrone, R., Butovskaya, E., Daelemans, D., Palù, G., Pannecouque, C., and Richter, S. N. (2014). Anti-HIV-1 activity of the G-quadruplex ligand BRACO-19. *J. Antimicrob. Chemother.* 69, 3248–3258. doi:10.1093/jac/dku280.

Pevzner, P., and Shamir, R. (2011). *Bioinformatics for biologists*.

Pfeiffer, J. K., and Kirkegaard, K. (2006). Bottleneck-mediated quasispecies restriction during spread of an RNA virus from inoculation site to brain. *Proc Natl Acad Sci USA* 103, 5520–5525. doi:10.1073/pnas.0600834103.

Picó, B., Díez, M. J., and Nuez, F. (1996). Viral diseases causing the greatest economic losses to the tomato crop. II. The Tomato yellow leaf curl virus — a review. *Scientia Horticulturae* 67, 151–196.

Piekna-Przybylska, D., Sullivan, M. A., Sharma, G., and Bambara, R. A. (2014). U3 region in the HIV-1 genome adopts a G-quadruplex structure in its RNA and DNA sequence. *Biochemistry* 53, 2581–2593. doi:10.1021/bi4016692.

Pilartz, M., and Jeske, H. (1992). Abutilon mosaic geminivirus double-stranded DNA is packed into minichromosomes. *Virology* 189, 800–802. doi:10.1016/0042-

6822(92)90610-2.

Pita, J. S., de Miranda, J. R., Schneider, W. L., and Roossinck, M. J. (2007). Environment determines fidelity for an RNA virus replicase. *Journal of Virology* 81, 9072–9077. doi:10.1128/JVI.00587-07.

Pita, J. S., Fondong, V. N., Sangaré, A., Otim-Nape, G. W., Ogwal, S., and Fauquet, C. M. (2001). Recombination, pseudorecombination and synergism of geminiviruses are determinant keys to the epidemic of severe cassava mosaic disease in Uganda. *Journal of General Virology* 82, 655–665. doi:10.1099/0022-1317-82-3-655.

Poon, A. F. Y., McGovern, R. A., Mo, T., Knapp, D. J. H. F., Brenner, B., Routy, J.-P., et al. (2011). Dates of HIV infection can be estimated for seroprevalent patients by coalescent analysis of serial next-generation sequencing data. *AIDS* 25, 2019–2026. doi:10.1097/QAD.0b013e32834b643c.

Quail, M. A., Smith, M., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., et al. (2012). A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* 13, 341. doi:10.1186/1471-2164-13-341.

Quiñones-Mateu, M. E., Avila, S., Reyes-Teran, G., and Martinez, M. A. (2014). Deep sequencing: Becoming a critical tool in clinical virology. *Journal of Clinical Virology* 61, 9–19. doi:10.1016/j.jcv.2014.06.013.

Radford, A. D., Chapman, D., Dixon, L., Chantrey, J., Darby, A. C., and Hall, N. (2012). Application of next-generation sequencing technologies in virology. *J. Gen. Virol.* 93, 1853–1868. doi:10.1099/vir.0.043182-0.

Ratner, L., Haseltine, W., Patarca, R., Livak, K. J., Starcich, B., Josephs, S. F., et al. (1985). Complete nucleotide sequence of the AIDS virus, HTLV-III. *Nature* 313, 277–284.

Rawson, J. M. O., Daly, M. B., Xie, J., Clouser, C. L., Landman, S. R., Reilly, C. S., et al. (2016). 5-Azacytidine Enhances the Mutagenesis of HIV-1 by Reduction to 5-Aza-2'-Deoxycytidine. *Antimicrob. Agents Chemother.* 60, 2318–2325. doi:10.1128/AAC.03084-15.

Recordon-Pinson, P., Raymond, S., Bellecave, P., Marcelin, A. G., Soulie, C., Descamps, D., et al. (2013). HIV-1 dynamics and coreceptor usage in Maraviroc-treated patients with ongoing replication. *Antimicrob. Agents Chemother.* 57, 930–935. doi:10.1128/AAC.02159-12.

Renzette, N., Bhattacharjee, B., Jensen, J. D., Gibson, L., and Kowalik, T. F. (2011). Extensive Genome-Wide Variability of Human Cytomegalovirus in Congenitally Infected Infants. *PLoS Pathog* 7, e1001344. doi:10.1371/journal.ppat.1001344.

Rest, J. S., and Mindell, D. P. (2003). SARS associated coronavirus has a recombinant polymerase and coronaviruses have a history of host-shifting. *Infection, Genetics and Evolution* 3, 219–225. doi:10.1016/j.meegid.2003.08.001.

Rey, L., Sidorova, J. M., Puget, N., Boudsocq, F., Biard, D. S. F., Monnat, R. J., et al. (2009). Human DNA polymerase eta is required for common fragile site stability during unperturbed DNA replication. *Mol. Cell. Biol.* 29, 3344–3354. doi:10.1128/MCB.00115-09.

Richter, K. S., and Jeske, H. (2015). KU80, a key factor for non-homologous end-joining, retards geminivirus multiplication. *J. Gen. Virol.* 96, 2913–2918. doi:10.1099/jgv.0.000224.

Richter, K. S., Götz, M., Winter, S., and Jeske, H. (2016a). The contribution of translesion synthesis polymerases on geminiviral replication. *Virology* 488, 137–148. doi:10.1016/j.virol.2015.10.027.

Richter, K. S., Götz, M., Winter, S., and Jeske, H. (2016b). The contribution of translesion synthesis polymerases on geminiviral replication. *Virology* 488, 137–148. doi:10.1016/j.virol.2015.10.027.

Richter, K. S., Götz, M., Winter, S., and Jeske, H. (2016c). The contribution of translesion synthesis polymerases on geminiviral replication. *Virology* 488, 137–148. doi:10.1016/j.virol.2015.10.027.

Richter, K. S., Serra, H., White, C. I., and Jeske, H. (2016d). The recombination mediator RAD51D promotes geminiviral infection. *Virology* 493, 113–127. doi:10.1016/j.virol.2016.03.014.

Ritzenthaler, C., and Hofmann, C. (2007). “Tubule-Guided Movement of Plant Viruses,” in *Viral transport in plants* Plant Cell Monographs. (Berlin, Heidelberg: Springer Berlin Heidelberg), 63–83. doi:10.1007/7089_2006_105.

Rodríguez-Cerezo, E., and García-Arenal, F. (1989). Genetic heterogeneity of the RNA genome population of the plant virus U5-TMV. *Virology* 170, 418–423. doi:10.1016/0042-6822(89)90432-7.

Rodríguez-Negrete, E., Lozano-Durán, R., Piedra-Aguilera, A., Cruzado, L., Bejarano, E. R., and Castillo, A. G. (2013). Geminivirus Rep protein interferes with the plant DNA methylation machinery and suppresses transcriptional gene silencing. *New Phytol.* 199, 464–475. doi:10.1111/nph.12286.

Rojas, A., Kvarnheden, A., Marcenaro, D., and Valkonen, J. P. T. (2005). Sequence characterization of tomato leaf curl Sinaloa virus and tomato severe leaf curl virus: phylogeny of New World begomoviruses and detection of recombination. *Arch Virol* 150, 1281–1299. doi:10.1007/s00705-005-0509-x.

Rojas, M. R., Jiang, H., Salati, R., Xoconostle-Cázares, B., Sudarshana, M. R., Lucas, W. J., et al. (2001). Functional Analysis of Proteins Involved in Movement of the Monopartite Begomovirus, Tomato Yellow Leaf Curl Virus. *Virology* 291, 110–125. doi:10.1006/viro.2001.1194.

Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlén, M., and Nyrén, P. (1996).

Real-time DNA sequencing using detection of pyrophosphate release. *Anal. Biochem.* 242, 84–89. doi:10.1006/abio.1996.0432.

Roossinck, M. J. (1997). Mechanisms of plant virus evolution. *Annu Rev Phytopathol* 35, 191–209. doi:10.1146/annurev.phyto.35.1.191.

Roossinck, M. J. (2002). Evolutionary history of Cucumber mosaic virus deduced by phylogenetic analyses. *Journal of Virology* 76, 3382–3387. doi:10.1128/JVI.76.7.3382-3387.2002.

Roossinck, M. J. (2005). Symbiosis versus competition in plant virus evolution. *Nat Rev Micro* 3, 917–924. doi:10.1038/nrmicro1285.

Ross, A.-L., and Sale, J. E. (2006). The catalytic activity of REV1 is employed during immunoglobulin gene diversification in DT40. *Mol. Immunol.* 43, 1587–1594. doi:10.1016/j.molimm.2005.09.017.

Rothberg, J. M., Hinz, W., Rearick, T. M., Schultz, J., Mileski, W., Davey, M., et al. (2011). An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 475, 348–352. doi:10.1038/nature10242.

Routh, A., and Johnson, J. E. (2014). Discovery of functional genomic motifs in viruses with ViReMa—a Virus Recombination Mapper—for analysis of next-generation sequencing data. *Nucleic Acids Research* 42, e11–e11. doi:10.1093/nar/gkt916.

Rozera, G., Abbate, I., Bruselles, A., Vlassi, C., D'Offizi, G., Narciso, P., et al. (2009). Massively parallel pyrosequencing highlights minority variants in the HIV-1 env quasispecies deriving from lymphomonocyte sub-populations. *Retrovirology* 6, 15. doi:10.1186/1742-4690-6-15.

Rubio, L., Ayllón, M. A., Kong, P., Fernández, A., Polek, M., Guerri, J., et al. (2001). Genetic variation of Citrus tristeza virus isolates from California and Spain: evidence for mixed infections and recombination. *Journal of Virology* 75, 8054–8062. doi:10.1128/JVI.75.17.8054-8062.2001.

Rudd, S. G., Bianchi, J., and Doherty, A. J. (2014). PrimPol—A new polymerase on the block. *Molecular & Cellular Oncology* 1, e960754. doi:10.4161/23723548.2014.960754.

Ruiz-Jarabo, C. M., Arias, A., Baranowski, E., Escarmís, C., and Domingo, E. (2000). Memory in viral quasispecies. *Journal of Virology* 74, 3543–3547.

Ruiz-Jarabo, C. M., Ly, C., Domingo, E., and Torre, J. C. de L. (2003). Lethal mutagenesis of the prototypic arenavirus lymphocytic choriomeningitis virus (LCMV). *Virology* 308, 37–47. doi:10.1016/S0042-6822(02)00046-6.

Ruiz-Medrano, R., Guevara-González, R. G., Argüello-Astorga, G. R., Monsalve-Fonnegra, Z., Herrera-Estrella, L. R., and Rivera-Bustamante, R. F. (1999). Identification of a sequence element involved in AC2-mediated transactivation of the pepper huasteco

virus coat protein gene. *Virology* 253, 162–169. doi:10.1006/viro.1998.9484.

Sacristán, S., Malpica, J. M., Fraile, A., and García-Arenal, F. (2003). Estimation of population bottlenecks during systemic movement of tobacco mosaic virus in tobacco plants. *Journal of Virology* 77, 9906–9911. doi:10.1128/JVI.77.18.9906-9911.2003.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 74, 5463–5467.

Sanjuán, R., Agudelo-Romero, P., and Elena, S. F. (2009). Upper-limit mutation rate estimation for a plant RNA virus. *Biol. Lett.* 5, 394–396. doi:10.1098/rsbl.2008.0762.

Sanjuán, R., Nebot, M. R., Chirico, N., Mansky, L. M., and Belshaw, R. (2010). Viral mutation rates. *Journal of Virology* 84, 9733–9748. doi:10.1128/JVI.00694-10.

Santiago, M. J., Ruiz-Rubio, M., Dio, L. D., González-Reyes, J. A., and Alejandre-Durán, E. (2008). Ubiquitous expression of two translesion synthesis DNA polymerase genes in Arabidopsis. *Planta* 227, 1269–1277. doi:10.1007/s00425-008-0698-0.

Sanz, A. I., Fraile, A., García-Arenal, F., Zhou, X., Robinson, D. J., Khalid, S., et al. (2000). Multiple infection, recombination and genome relationships among begomovirus isolates found in cotton and other plants in Pakistan. *Journal of General Virology* 81, 1839–1849. doi:10.1099/0022-1317-81-7-1839.

Sarkies, P., Reams, C., Simpson, L. J., and Sale, J. E. (2010). Epigenetic instability due to defective replication of structured DNA. *Mol. Cell* 40, 703–713. doi:10.1016/j.molcel.2010.11.009.

Saunders, K., Salim, N., Mali, V. R., Malathi, V. G., Briddon, R., Markham, P. G., et al. (2002). Characterisation of Sri Lankan cassava mosaic virus and Indian cassava mosaic virus: evidence for acquisition of a DNA B component by a monopartite begomovirus. *Virology* 293, 63–74. doi:10.1006/viro.2001.1251.

Sánchez-Campos, S., Díaz, J. A., Monci, F., Bejarano, E. R., Reina, J., Navas-Castillo, J., et al. (2002). High Genetic Stability of the Begomovirus Tomato yellow leaf curl Sardinia virus in Southern Spain Over an 8-Year Period. *Phytopathology* 92, 842–849. doi:10.1094/PHYTO.2002.92.8.842.

Sánchez-Campos, S., Navas-Castillo, J., Camero, R., Soria, C., Díaz, J. A., and Moriones, E. (1999). Displacement of Tomato Yellow Leaf Curl Virus (TYLCV)-Sr by TYLCV-Is in Tomato Epidemics in Spain. *Phytopathology* 89, 1038–1043. doi:10.1094/PHYTO.1999.89.11.1038.

Sánchez-Durán, M. A., Dallas, M. B., Ascencio-Ibañez, J. T., Reyes, M. I., Arroyo-Mateos, M., Ruiz-Albert, J., et al. (2011). Interaction between geminivirus replication protein and the SUMO-conjugating enzyme is required for viral infection. *Journal of Virology* 85, 9789–9800. doi:10.1128/JVI.02566-10.

Schiavone, D., Jozwiakowski, S. K., Romanello, M., Guilbaud, G., Guillian, T. A.,

Bailey, L. J., et al. (2016). PrimPol Is Required for Replicative Tolerance of G Quadruplexes in Vertebrate Cells. *Mol. Cell* 61, 161–169. doi:10.1016/j.molcel.2015.10.038.

Schneider, W. L., and Roossinck, M. J. (2000). Evolutionarily related Sindbis-like plant viruses maintain different levels of population diversity in a common host. *Journal of Virology* 74, 3130–3134.

Schneider, W. L., and Roossinck, M. J. (2001). Genetic diversity in RNA virus quasispecies is controlled by host-virus interactions. *Journal of Virology* 75, 6566–6571. doi:10.1128/JVI.75.14.6566-6571.2001.

Selth, L. A., Dogra, S. C., Rasheed, M. S., Healy, H., Randles, J. W., and Rezaian, M. A. (2005). A NAC domain protein interacts with tomato leaf curl virus replication accessory protein and enhances viral replication. *The Plant Cell* 17, 311–325. doi:10.1105/tpc.104.027235.

Selth, L. A., Randles, J. W., and Rezaian, M. A. (2004). Host responses to transient expression of individual genes encoded by Tomato leaf curl virus. *MPMI* 17, 27–33. doi:10.1094/MPMI.2004.17.1.27.

Settlage, S. B., See, R. G., and Hanley-Bowdoin, L. (2005). Geminivirus C3 protein: replication enhancement and protein interactions. *Journal of Virology* 79, 9885–9895. doi:10.1128/JVI.79.15.9885-9895.2005.

Shao, W., Boltz, V. F., Spindler, J. E., Kearney, M. F., Maldarelli, F., Mellors, J. W., et al. (2013). Analysis of 454 sequencing error rate, error sources, and artifact recombination for detection of Low-frequency drug resistance mutations in HIV-1 DNA. *Retrovirology* 10, 18. doi:10.1186/1742-4690-10-18.

Sharma, S., Hicks, J. K., Chute, C. L., Brennan, J. R., Ahn, J.-Y., Glover, T. W., et al. (2012). REV1 and polymerase ζ facilitate homologous recombination repair. *Nucleic Acids Research* 40, 682–691. doi:10.1093/nar/gkr769.

Shaw, J. G., Plaskitt, K. A., and Wilson, T. M. (1986). Evidence that tobacco mosaic virus particles disassemble contrtranslationally in vivo. *Virology* 148, 326–336.

Sierra, M., Airaksinen, A., González- López, C., Agudo, R., Arias, A., and Domingo, E. (2007). Foot-and-mouth disease virus mutant with decreased sensitivity to ribavirin: implications for error catastrophe. *Journal of Virology* 81, 2012–2024. doi:10.1128/JVI.01606-06.

Sierra, S., Dávila, M., Lowenstein, P. R., and Domingo, E. (2000). Response of foot-and-mouth disease virus to increased mutagenesis: influence of viral load and fitness in loss of infectivity. *Journal of Virology* 74, 8316–8323.

Silverstein, T. D., Johnson, R. E., Jain, R., Prakash, L., Prakash, S., and Aggarwal, A. K. (2010). Structural basis for the suppression of skin cancers by DNA polymerase ϵ . *Nature* 465, 1039–1043. doi:10.1038/nature09104.

Simmons, H. E., Dunham, J. P., Stack, J. C., Dickins, B. J. A., Pagán, I., Holmes, E. C., et al. (2012). Deep sequencing reveals persistence of intra- and inter-host genetic diversity in natural and greenhouse populations of zucchini yellow mosaic virus. *J. Gen. Virol.* 93, 1831–1840. doi:10.1099/vir.0.042622-0.

Singh, D. K., Islam, M. N., Choudhury, N. R., Karjee, S., and Mukherjee, S. K. (2007). The 32 kDa subunit of replication protein A (RPA) participates in the DNA replication of Mung bean yellow mosaic India virus (MYMIV) by interacting with the viral Rep protein. *Nucleic Acids Research* 35, 755–770. doi:10.1093/nar/gkl1088.

Sobrino, F., Palma, E. L., Beck, E., Dávila, M., la Torre, de, J. C., Negro, P., et al. (1986). Fixation of mutations in the viral genome during an outbreak of foot-and-mouth disease: heterogeneity and rate variations. *Gene* 50, 149–159.

Spandole, S., Cimponeriu, D., Berca, L. M., and Mihăescu, G. (2015). Human anelloviruses: an update of molecular, epidemiological and clinical aspects. *Arch Virol* 160, 893–908. doi:10.1007/s00705-015-2363-9.

Stanley, J. (2005). *Geminiviridae*, p 301–326. Virus taxonomy. Eighth report of the International

Stanley, W. M. (1935). Isolation of a crystalline protein possessing the properties of tobacco mosaic virus. *Science*.

Stasolla, C., Katahira, R., Thorpe, T. A., and Ashihara, H. (2003). Purine and pyrimidine nucleotide metabolism in higher plants. *J. Plant Physiol.* 160, 1271–1295. doi:10.1078/0176-1617-01169.

Sulzinski, M. A., Gabard, K. A., Palukaitis, P., and Zaitlin, M. (1985). Replication of tobacco mosaic virus. VIII. Characterization of a third subgenomic TMV RNA. *Virology* 145, 132–140.

Sunter, G., Hartitz, M. D., Hormuzdi, S. G., Brough, C. L., and Bisaro, D. M. (1990). Genetic analysis of tomato golden mosaic virus: ORF AL2 is required for coat protein accumulation while ORF AL3 is necessary for efficient DNA replication. *Virology* 179, 69–77.

Špak, J., Votruba, I., Pavingerová, D., Holý, A., Špaková, V., and Petrzik, K. (2011). Antiviral activity of tenofovir against Cauliflower mosaic virus and its metabolism in Brassica pekinensis plants. *Antiviral Research* 92, 378–381. doi:10.1016/j.antiviral.2011.08.014.

Takahashi, S., Sakamoto, A. N., Tanaka, A., and Shimizu, K. (2007). AtREV1, a Y-Family DNA Polymerase in Arabidopsis, Has Deoxynucleotidyl Transferase Activity in Vitro. *PLANT PHYSIOLOGY* 145, 1052–1060. doi:10.1104/pp.107.101980.

Takahashi, S., Sakamoto, A., Sato, S., Kato, T., Tabata, S., and Tanaka, A. (2005). Roles of Arabidopsis AtREV1 and AtREV7 in translesion synthesis. *PLANT PHYSIOLOGY* 138, 870–881. doi:10.1104/pp.105.060236.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi:10.1093/molbev/mst197.

Tan, Z., Wada, Y., Chen, J., and Ohshima, K. (2004). Inter- and intralinear recombinants are common in natural populations of Turnip mosaic virus. *Journal of General Virology* 85, 2683–2696. doi:10.1099/vir.0.80124-0.

Tanguay, R. L., and Gallie, D. R. (1996). Translational efficiency is regulated by the length of the 3' untranslated region. *Mol. Cell. Biol.* 16, 146–156.

Teng, M. N., Oldstone, M. B., and la Torre, de, J. C. (1996). Suppression of lymphocytic choriomeningitis virus--induced growth hormone deficiency syndrome by disease-negative virus variants. *Virology* 223, 113–119. doi:10.1006/viro.1996.0460.

Tissier, A., Kannouche, P., Reck, M.-P., Lehmann, A. R., Fuchs, R. P. P., and Cordonnier, A. (2004). Co-localization in replication foci and interaction of human Y-family members, DNA polymerase pol eta and REV1 protein. *DNA Repair (Amst.)* 3, 1503–1514. doi:10.1016/j.dnarep.2004.06.015.

Tlučková, K., Marušič, M., Tóthová, P., Bauer, L., Šket, P., Plavec, J., et al. (2013). Human Papillomavirus G-Quadruplexes. *Biochemistry* 52, 7207–7216. doi:10.1021/bi400897g.

Töpfer, A., Marschall, T., Bull, R. A., Luciani, F., Schönhuth, A., and Beerenwinkel, N. (2014). Viral Quasispecies Assembly via Maximal Clique Enumeration. *PLoS Comput Biol* 10, e1003515–10. doi:10.1371/journal.pcbi.1003515.

Trincao, J., Johnson, R. E., Escalante, C. R., Prakash, S., Prakash, L., and Aggarwal, A. K. (2001). Structure of the catalytic core of *S. cerevisiae* DNA polymerase eta: implications for translesion DNA synthesis. *Mol. Cell* 8, 417–426.

Tromas, N., and Elena, S. F. (2010). The rate and spectrum of spontaneous mutations in a plant RNA virus. *Genetics* 185, 983–989. doi:10.1534/genetics.110.115915.

Tsibris, A. M. N., Korber, B., Arnaout, R., Russ, C., Lo, C.-C., Leitner, T., et al. (2009). Quantitative deep sequencing reveals dynamic HIV-1 escape and large population shifts during CCR5 antagonist therapy in vivo. *PLoS ONE* 4, e5683. doi:10.1371/journal.pone.0005683.

Tugume, A. K., Mukasa, S. B., Kalkkinen, N., and Valkonen, J. P. T. (2010). Recombination and selection pressure in the ipomovirus sweet potato mild mottle virus (Potyviridae) in wild species and cultivated sweetpotato in the centre of evolution in East Africa. *J. Gen. Virol.* 91, 1092–1108. doi:10.1099/vir.0.016089-0.

Ueki, S., and Citovsky, V. (2011). To gate, or not to gate: regulatory mechanisms for intercellular protein transport and virus movement in plants. *Mol Plant* 4, 782–793. doi:10.1093/mp/ssr060.

Valouev, A., Ichikawa, J., Tonthat, T., Stuart, J., Ranade, S., Peckham, H., et al. (2008). A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning. *Genome Res.* 18, 1051–1063. doi:10.1101/gr.076463.108.

van der Walt, E., Martin, D. P., Varsani, A., Polston, J. E., and Rybicki, E. P. (2008). Experimental observations of rapid Maize streak virus evolution reveal a strand-specific nucleotide substitution bias. *Virology Journal* 5, 104. doi:10.1186/1743-422X-5-104.

van Kregten, M., and Tijsterman, M. (2014). The repair of G-quadruplex-induced DNA damage. *Experimental Cell Research* 329, 178–183. doi:10.1016/j.yexcr.2014.08.038.

Van Loy, T., Thys, K., Ryschkewitsch, C., Lagatie, O., Monaco, M. C., Major, E. O., et al. (2015). JC Virus Quasispecies Analysis Reveals a Complex Viral Population Underlying Progressive Multifocal Leukoencephalopathy and Supports Viral Dissemination via the Hematogenous Route. *Journal of Virology* 89, 1340–1347. doi:10.1128/JVI.02565-14.

Van Loy, T., Thys, K., Tritsmans, L., and Stuyver, L. J. (2013). Quasispecies analysis of JC virus DNA present in urine of healthy subjects. *PLoS ONE* 8, e70950. doi:10.1371/journal.pone.0070950.

Vanitharani, R., Chellappan, P., Pita, J. S., and Fauquet, C. M. (2004). Differential roles of AC2 and AC4 of cassava geminiviruses in mediating synergism and suppression of posttranscriptional gene silencing. *Journal of Virology* 78, 9487–9498. doi:10.1128/JVI.78.17.9487-9498.2004.

Varghese, V., Shahriar, R., Rhee, S.-Y., Liu, T., Simen, B. B., Egholm, M., et al. (2009). Minority variants associated with transmitted and acquired HIV-1 nonnucleoside reverse transcriptase inhibitor resistance: implications for the use of second-generation nonnucleoside reverse transcriptase inhibitors. *J. Acquir. Immune Defic. Syndr.* 52, 309–315. doi:10.1097/QAI.0b013e3181bca669.

Varsani, A., Navas-Castillo, J., Moriones, E., Hernández-Zepeda, C., Idris, A., Brown, J. K., et al. (2014). Establishment of three new genera in the family Geminiviridae: Becurtovirus, Eragrovirus and Turncurtovirus. *Arch Virol* 159, 2193–2203. doi:10.1007/s00705-014-2050-2.

Vignuzzi, M., Stone, J. K., Arnold, J. J., Cameron, C. E., and Andino, R. (2006). Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature* 439, 344–348. doi:10.1038/nature04388.

Viguera, E., Canceill, D., and Ehrlich, S. D. (2001). Replication slippage involves DNA polymerase pausing and dissociation. *EMBO J.* 20, 2587–2595. doi:10.1093/emboj/20.10.2587.

Viral Genome Methylation as an Epigenetic Defense against Geminiviruses (2008). Viral Genome Methylation as an Epigenetic Defense against Geminiviruses. *Journal of Virology* 82, 8997–9007. doi:10.1128/JVI.00719-08.

Voinnet, O., Pinto, Y. M., and Baulcombe, D. C. (1999). Suppression of gene silencing: a general strategy used by diverse DNA and RNA viruses of plants. *Proc Natl Acad Sci USA* 96, 14147–14152. doi:10.1073/pnas.96.24.14147.

Volkenstein, P. M. V. (1994). "Informational Aspects of Evolution," in *Physical approaches to biological evolution* (Berlin, Heidelberg: Springer Berlin Heidelberg), 335–366. doi:10.1007/978-3-642-78788-1_10.

Wald, A., and Wolfowitz, J. (1943). An exact test for randomness in the non-parametric case based on serial correlation. *The Annals of Mathematical Statistics*.

Wang, E., and Wagner, G. J. (2003). Elucidation of the functions of genes central to diterpene metabolism in tobacco trichomes using posttranscriptional gene silencing. *Planta* 216, 686–691. doi:10.1007/s00425-002-0904-4.

Wang, Z., Huang, M., Ma, X., Li, H., Tang, T., and Guo, C. (2016). REV1 promotes PCNA monoubiquitination through interacting with ubiquitinated RAD18. *J. Cell. Sci.* 129, 1223–1233. doi:10.1242/jcs.179408.

Ward, B. M., Medville, R., Lazarowitz, S. G., and Turgeon, R. (1997). The geminivirus BL1 movement protein is associated with endoplasmic reticulum-derived tubules in developing phloem cells. *Journal of Virology* 71, 3726–3733.

Wartig, L., Kheyr-Pour, A., Noris, E., De Kouchkovsky, F., Jouanneau, F., Gronenborn, B., et al. (1997). Genetic analysis of the monopartite tomato yellow leaf curl geminivirus: roles of V1, V2, and C2 ORFs in viral pathogenesis. *Virology* 228, 132–140. doi:10.1006/viro.1996.8406.

Washington, M. T., Johnson, R. E., Prakash, L., and Prakash, S. (2002). Human DINB1-encoded DNA polymerase kappa is a promiscuous extender of mispaired primer termini. *Proc Natl Acad Sci USA* 99, 1910–1914. doi:10.1073/pnas.032594399.

Waters, L. S., Minesinger, B. K., Wiltout, M. E., D'Souza, S., Woodruff, R. V., and Walker, G. C. (2009). Eukaryotic Translesion Polymerases and Their Roles and Regulation in DNA Damage Tolerance. *Microbiology and Molecular Biology Reviews* 73, 134–154. doi:10.1128/MMBR.00034-08.

Whitehurst, C. B., Vaziri, C., Shackelford, J., and Pagano, J. S. (2012). Epstein-Barr virus BPLF1 deubiquitinates PCNA and attenuates polymerase η recruitment to DNA damage sites. *Journal of Virology* 86, 8097–8106. doi:10.1128/JVI.00588-12.

Wickramasinghe, C. M., Arzouk, H., Frey, A., Maiter, A., and Sale, J. E. (2015). Contributions of the specialised DNA polymerases to replication of structured DNA. *DNA Repair (Amst.)* 29, 83–90. doi:10.1016/j.dnarep.2015.01.004.

Wilson, T. M. (1984). Cotranslational disassembly of tobacco mosaic virus in vitro. *Virology* 137, 255–265.

Wolf, S., Deom, C. M., Beachy, R. N., and Lucas, W. J. (1989). Movement protein of

tobacco mosaic virus modifies plasmodesmatal size exclusion limit. *Science* 246, 377–379.

Wu, X., and Shaw, J. G. (1997). Evidence that a viral replicase protein is involved in the disassembly of tobacco mosaic virus particles in vivo. *Virology* 239, 426–434. doi:10.1006/viro.1997.8870.

Wu, X., Xu, Z., and Shaw, J. G. (1994). Uncoating of tobacco mosaic virus RNA in protoplasts. *Virology* 200, 256–262. doi:10.1006/viro.1994.1183.

Wu, X., Zhou, T., Zhu, J., Zhang, B., Georgiev, I., Wang, C., et al. (2011). Focused evolution of HIV-1 neutralizing antibodies revealed by structures and deep sequencing. *Science* 333, 1593–1602. doi:10.1126/science.1207532.

Xiao, M., Prabakaran, P., Chen, W., Kessing, B., and Dimitrov, D. S. (2013). Deep sequencing and Circos analyses of antibody libraries reveal antigen-driven selection of Ig VH genes during HIV-1 infection. *Exp. Mol. Pathol.* 95, 357–363. doi:10.1016/j.yexmp.2013.10.004.

Xu, J., Zhang, Y., and Pun, N. (2013). Mitochondrial recombination in natural populations of the button mushroom *Agaricus bisporus*. *Fungal Genet. Biol.* 55, 92–97. doi:10.1016/j.fgb.2012.09.004.

Yadav, C. B., Bhareti, P., Muthamilarasan, M., Mukherjee, M., Khan, Y., Rathi, P., et al. (2015). Genome-wide SNP identification and characterization in two soybean cultivars with contrasting Mungbean Yellow Mosaic India Virus disease resistance traits. *PLoS ONE* 10, e0123897. doi:10.1371/journal.pone.0123897.

Yamada, K., Noguchi, K., Komeno, T., Furuta, Y., and Nishizono, A. (2016). Efficacy of Favipiravir (T-705) in Rabies Postexposure Prophylaxis. *J. Infect. Dis.* 213, 1253–1261. doi:10.1093/infdis/jiv586.

Yin, L., Liu, L., Sun, Y., Hou, W., Lowe, A. C., Gardner, B. P., et al. (2012). High-resolution deep sequencing reveals biodiversity, population structure, and persistence of HIV-1 quasispecies within host ecosystems. *Retrovirology* 9, 1–1. doi:10.1186/1742-4690-9-108.

Yu, X.-Q., Jia, J.-L., Zhang, C.-L., Li, X.-D., and Wang, Y.-J. (2010). Phylogenetic analyses of an isolate obtained from potato in 1985 revealed potato virus X was introduced to China via multiple events. *Virus Genes* 40, 447–451. doi:10.1007/s11262-010-0468-5.

Zbawicka, M., Wenne, R., and Burzyński, A. (2014). Mitogenomics of recombinant mitochondrial genomes of Baltic Sea *Mytilus* mussels. *Mol Genet Genomics* 289, 1275–1287. doi:10.1007/s00438-014-0888-3.

Zeng, X., Winter, D. B., Kasmer, C., Kraemer, K. H., Lehmann, A. R., and Gearhart, P. J. (2001). DNA polymerase [eta] is an A-T mutator in somatic hypermutation of immunoglobulin variable genes. *Nature Immunology* 2, 537–541. doi:10.1038/88740.

Zhang, W., Chen, J., Yang, Y., Tang, Y., Shang, J., and Shen, B. (2011). A Practical Comparison of De Novo Genome Assembly Software Tools for Next-Generation Sequencing Technologies. *PLoS ONE* 6, e17915. doi:10.1371/journal.pone.0017915.

Zhang, Y., Wu, X., Rechkoblit, O., Geacintov, N. E., Taylor, J.-S., and Wang, Z. (2002). Response of human REV1 to different DNA damage: preferential dCMP insertion opposite the lesion. *Nucleic Acids Research* 30, 1630–1638.

Zhang, Y., Yuan, F., Wu, X., Rechkoblit, O., Taylor, J. S., Geacintov, N. E., et al. (2000). Error-prone lesion bypass by human DNA polymerase eta. *Nucleic Acids Research* 28, 4717–4724.

Zhang, Y., Yuan, T., Li, J., Zhang, Y., Xu, J., Shao, Y., et al. (2013). The potential of the human immune system to develop broadly neutralizing HIV-1 antibodies: implications for vaccine development. *AIDS* 27, 2529–2539. doi:10.1097/QAD.000000000000015.

Zhou, X., Liu, Y., Robinson, D. J., and Harrison, B. D. (1998). Four DNA-A variants among Pakistani isolates of cotton leaf curl virus and their affinities to DNA-A of geminivirus isolates from okra. *Journal of General Virology* 79 (Pt 4), 915–923. doi:10.1099/0022-1317-79-4-915.

Ziv, O., Geacintov, N., Nakajima, S., Yasui, A., and Livneh, Z. (2009). DNA polymerase zeta cooperates with polymerases kappa and iota in translesion DNA synthesis across pyrimidine photodimers in cells from XPV patients. *Proc. Natl. Acad. Sci. U.S.A.* 106, 11552–11557. doi:10.1073/pnas.0812548106.

Zlatanou, A., Despras, E., Braz-Petta, T., Boubakour-Azzouz, I., Pouvelle, C., Stewart, G. S., et al. (2011). The hMsh2-hMsh6 Complex Acts in Concert with Monoubiquitinated PCNA and Pol η in Response to Oxidative DNA Damage in Human Cells. *Mol. Cell* 43, 649–662. doi:10.1016/j.molcel.2011.06.023.

Zrachya, A., Kumar, P. P., Ramakrishnan, U., Levy, Y., Loyter, A., Arazi, T., et al. (2007). Production of siRNA targeted against TYLCV coat protein transcripts leads to silencing of its expression and resistance to the virus. *Transgenic Res* 16, 385–398. doi:10.1007/s11248-006-9042-2.